

# *eContentplus*

## **Best Practice Network**

*Action 5.1 Best Practice Networks for interoperability of digital libraries*

### **BHL-Europe**

Biodiversity Heritage Library for Europe:

Making biodiversity literature in European languages globally available.

#### *Description of Work*

##### **Proposal abstract**

The lack of access to the published biodiversity literature is a major obstacle to efficient research and a broad range of other applications, including education, biodiversity conservation, protected area management, disease control, and maintenance of diverse ecosystems services. This literature also has cultural importance as a resource for the study of the history of science, art and other non-science applications. Currently, a large number of small projects are digitising biodiversity material in numerous institutions across the EU to make access more open, but the corpus will still be seriously fragmented. These projects do not use common standards or interfaces and are not interoperable. In alignment with the EC i2010 initiative, BHL-Europe aims to make the biodiversity knowledge available to everybody who is interested by improving the interoperability of European biodiversity digital libraries.

BHL-Europe will review and test different approaches for such libraries based on the experiences of the partners involved in the project. The consortium will establish a best practice approach and promote the adoption of standards and specifications for the large-scale implementation in a real-life context. BHL-Europe will provide a multilingual access point for search and retrieval of digital content through EUROPEANA. In addition, it will provide a robust multilingual portal with sophisticated search tools to facilitate the search for taxa-specific biodiversity information. The project will also develop operational strategies and processes for long-term preservation and sustainability of the data produced by national biodiversity digitisation programmes. BHL-Europe will generate activities to raise awareness and to ensure that the project outputs are known and used by the target users and that the proposed approach directly addresses user needs. BHL-Europe experience and best practice will be shared with the wider digital library community.

## Table of contents

1	RATIONALE AND OBJECTIVES .....	3
1.1	<i>Description of the issue and proposed solution</i> .....	3
1.2	<i>Expected results</i> .....	6
1.3	<i>List of participants</i> .....	6
2	CONTRIBUTION TO ECONTENTPLUS PROGRAMME OBJECTIVES .....	10
3	EUROPEAN DIMENSION .....	11
4	CONTENT .....	11
4.1	<i>IPR issues</i> .....	16
4.2	<i>Multilingual and/or multicultural aspects</i> .....	16
5	IMPACT .....	17
5.1	<i>Analysis of demand</i> .....	17
5.2	<i>Target users and their needs</i> .....	18
5.3	<i>Critical Mass</i> .....	19
5.4	<i>Added Value</i> .....	20
5.5	<i>Success indicators</i> .....	20
5.6	<i>Dissemination and awareness</i> .....	20
6	NETWORKING .....	22
6.1	<i>Networking Capacity</i> .....	22
6.2	<i>Clustering Activities</i> .....	23
7	PROJECT WORK PLAN .....	24
7.1	<i>Introduction and general description</i> .....	24
7.2	<i>Work package overview</i> .....	27
7.3	<i>Work package description</i> .....	28
7.4	<i>Deliverables List</i> .....	36
7.5	<i>Project plan</i> .....	38
8	PROJECT MANAGEMENT .....	39
8.1	<i>Project Management Structure</i> .....	39
8.2	<i>Project Coordination (PCO)</i> .....	39
8.3	<i>Project Steering Group (PSG)</i> .....	40
8.4	<i>Project Management Group (PMG)</i> .....	40
8.5	<i>Technology Management Board (TMB)</i> .....	41
8.6	<i>Communications Working Group (CWG)</i> .....	41
8.7	<i>Work Packages (WP)</i> .....	41
8.8	<i>Means for communication, monitoring and assuring quality</i> .....	42
8.9	<i>Means for resolving conflicts and risk management</i> .....	42
9	DESCRIPTION OF CONSORTIUM AND KEY PERSONNEL .....	43
9.1	<i>Description of partners and key personnel</i> .....	43
10	APPENDICES .....	58
10.1	<i>Financing Plan</i> .....	58
10.2	<i>Other specific costs</i> .....	59
10.3	<i>List of events and meetings</i> .....	60
10.4	<i>Background and reference documents</i> .....	61

# 1 *Rationale and Objectives*

## 1.1 **Description of the issue and proposed solution**

### 1.1.1 **Background: the importance of biodiversity literature**

The libraries of the European natural history museums and botanical gardens collectively hold the majority of the world's published knowledge on the discovery and subsequent description of biological diversity. As yet this wealth of knowledge is only currently available to those few people who can gain direct access to these collections. The body of biodiversity knowledge is thus effectively withheld from use for a wide range of scientific applications, which include research, education, taxonomic study, biodiversity conservation, protected area management, disease control, and maintenance of diverse ecosystems services. Much of the early published literature is rare or has limited global distribution and is available in only a very few libraries. From a research perspective, these collections are of exceptional value because the domain of systematic biology depends – more than any other natural science – upon historic literature. The cited “half-life” (period of relevance) of natural history literature is longer than that of any other scientific domain and its “decay-rate” (rate at which it becomes irrelevant) is much slower than in other fields (cf. biotechnology). In order to positively identify a rare specimen, a working biologist may still have to consult a 100 year-old text, because that was the last time the organism was found, recorded, and described.

Once the collections of biodiversity literature are freely available on the Internet, this will be of great value to scientists, and also to a much wider public. Amateur naturalists (citizen scientists) who lack affiliation with major research institutions will be able to search, read, download, and print articles that were previously unavailable to them. Designers and artists will be able to use the detailed illustrations integral to many taxonomic works as motifs or design concepts in their work. Educators guiding students in how to do biological research will have access to a wealth of examples to incorporate into lesson plans and assignments. Historians and social scientists will have access to the stories and background information on the development of the natural sciences and to ever-developing scientific theories and understanding.

The greatest diversity of the biota exists in tropical and developing countries, yet the literature documenting this biodiversity is overwhelmingly held in a small number of European and North American libraries. Digitising this literature and making it freely available on the Internet is an act of significant knowledge transfer, thus helping the EU to achieve its commitments under Article 17 of the Convention on Biological Diversity (CBD).

In addition to the scientific value of this literature, the taxonomic literature is also part of our cultural heritage. Taxonomists study and describe the organisms and biodiversity of particular areas. These areas commonly are cultural landscapes or parts of cultural landscapes. Cultural landscapes “represent the combined works of nature and of man” (UNESCO World Heritage Convention II.A.47) and reflect the evolution of human society over time in relationship to its ecological context. Thus cultural landscapes are part of the cultural heritage according to Article 1 of the UNESCO World Heritage Convention. As part of this cultural landscape, information on nature and biodiversity will be combined with information on archaeology and ethnology through the lead project of the European Digital Library Foundation, EUROPEANA. Descriptions and documentations of natural phenomena, of plants, and of animals, should be considered as part of the European cultural heritage.

The work of Maria Sibylla Merian (1647-1717) is a good example of the integration of natural and cultural heritage: as a pioneer in art and science, she combined her talent for painting with her interest in natural history to produce exceptional paintings of natural objects and richly illustrated studies on flowers and caterpillars. Recent appreciation of her work in science and art is seen in two art exhibitions on her work in Amsterdam (Het Rembrandthuis) and Los Angeles (J Paul Getty Museum).

BHL-Europe directly addresses the overall aim of *eContentplus*, “to make digital content in Europe more accessible, usable and exploitable” and to do this in the context of key information sources for biodiversity. More specifically it focuses on **Action 5.1. Best Practice Networks for interoperability of digital libraries** with the objective of improving the interoperability of digital libraries held by 15 natural history museums and botanic gardens and two institutional archives, and libraries across 12 EU Member States, and will be progressively extendable. BHL Europe will establish standards-based interoperability between digitised documentation and the collections, ensuring the interoperability of the systems and multilingual and cross-cultural search and retrieval of the digital content, and facilitate user access through the common user interface of EUROPEANA.

### 1.1.2 EUROPEANA and the European Digital Library Foundation

EDLNet and other major projects within Europe are leading to the creation of EUROPEANA – a single cultural portal for digital materials for Europe. The European Digital Library Foundation has been set up to develop a sustainable and secure future for this portal and the underlying digital objects. EDLNet has taken the lead in establishing best practice and standards for the creation of digital objects and for the development of appropriate metadata, leading to a large-scale distributed portal model.

Currently, the majority of material in EUROPEANA is from the arts and humanities. However, the European Digital Library was set up to also include science. The biodiversity heritage literature is an excellent candidate for the first major corpus of science material to join EUROPEANA – the literature has both scientific value and major cultural importance. These biodiversity texts are the core to any study of the history of the development of scientific thought in Europe over the last three centuries.

### 1.1.3 Background to the global Biodiversity Heritage Library (BHL) Project

Since 2007, ten major biodiversity libraries have collaborated in digitising the biodiversity literature (main focus on English language literature) in an open access manner via the Biodiversity Heritage Library (BHL) project. Two European institutions are participating in the BHL project: Natural History Museum (London, UK) and Royal Botanic Gardens, Kew (Richmond, UK). Preliminary commitments have also been received from the Australian government and Chinese Academy of Sciences to digitise their own biodiversity literature. The global BHL project is predominantly funded by grants from large US foundations and the operating budgets of the member institutions.

The main aim of the BHL project is to make the biodiversity knowledge accessible on an open access Creative Commons basis to a wide spectrum of end users ranging from researchers, environmental agencies to members of the public with strong and effective outreach. Optical Character Recognition (OCR) techniques will be used to extract the relevant information from digitised objects. Taxonomic intelligence tools are used to overcome the issue of changing names, names using local languages and plurality of names for the same object (synonyms). A research scientist, student or member of the public, who has access to the Internet anywhere in the world, will be able to search for specific information in all of the literature relevant to biodiversity and transparently link the documentation to taxonomic, geographic, biographic, or other relevant databases.

It is essential that BHL should be a global partnership because, while many libraries have collected biodiversity materials, no single library holds the complete corpus of legacy literature. The partners' collections will represent a comprehensive assemblage of this literature. The BHL, as a community-based partnership, will provide a trusted grouping to negotiate with copyright owners. Within two years of the start of this project, the BHL will have made approximately 15 million digitised pages of literature available to support multiple biodiversity informatics initiatives and research. For the

first time in history, the core of our natural history museum and botanical garden libraries will be available to a truly global audience.

The BHL Portal is available at [www.biodiversitylibrary.org](http://www.biodiversitylibrary.org), and fuller details of the BHL program can be found in the Appendix, section 10.2.1.

#### 1.1.4 Issues addressed by BHL-Europe

The BHL is a global project and it is vital that Europe contributes its biodiversity literature to the project and that European users have access to the global BHL project. Some of the important English language literature of Europe is already part of the BHL. It is essential that the very significant amount of biodiversity literature held in other European languages – German, French, Dutch, Spanish, Italian, and others – is also integrated and becomes far more widely accessible to users. This requires a European effort to establish BHL-Europe as a dynamic component that will be both valuable in itself and also contributes to a global effort.

**BHL-Europe** will manage the acquisition, digitisation, and hosting of the material contained in European institutions and will significantly improve the interoperability of these currently disparate and developing European digital libraries. The material in each European nation will be made accessible through **EUROPEANA** (European Digital Library Foundation). **BHL-Europe** will manage the process by which each nation digitises its biodiversity material and ensure that this is done efficiently and effectively. The efficient coordination and management of ‘commitments to digitise’ is vital – some of the material is available in a number of different locations and institutions in Europe and duplication must be avoided.

**BHL-Europe** will review and test different approaches, standards and specifications for biodiversity digital libraries based on the experiences of the partners involved in the project. The consortium partners will agree on a best practice approach to be used for the large-scale implementation in this real-life context.

**BHL-Europe** will manage the relationship with the global BHL project and with the national and European partners. European experience in multilingual and multicultural database technology will add considerable value to the BHL portal by providing multilingual access. Our experience in distributed data management will be vital to the long-term sustainability of the whole BHL project.

**BHL-Europe** is a project that will mobilise funding in individual EU nations in order to undertake and complete essential scanning work. Several of the partners have already indicated that their government will promote national scanning initiatives once a large coordinating project like BHL-Europe is in place.

#### 1.1.5 Benefits of BHL-Europe

- 1) Enabling access to about 20 million pages of scientific literature on biodiversity to EUROPEANA will have a significant impact on the breadth and depth of European culture covered. Specifically, the action will:
  - begin the fulfilment of the scientific dimension of the EUROPEANA cultural arena
  - demonstrate the global importance of European scientific endeavour in biological sciences
  - provide tools and information for the study of the history and sociology of European science
  - provide access to culturally-important documents – from Darwin, Linnaeus, von Humboldt, Wallace, Cuvier, Merian, etc.
  - provide access to many beautiful and culturally-important images – botanical drawings, zoological drawings, watercolours, etc.
- 2) Providing access to the biodiversity literature (images and text) using a common global portal with integrated and sophisticated search tools will produce a number of long-term benefits for the European and global biology communities. These outcomes include:

- improving the efficiency of research in the biology domain
  - improving access to biodiversity information for non-museum biologists
  - repatriation of information about species in developing countries back to those countries via the Web
  - capacity building in the developing world
  - preservation of rare and fragile materials
- 3) The project mobilises 26 partners (22 of which are (potential) literature providers) from 12 EU countries (of which 3 are New Member States). It is an intersectoral consortium of museums, botanic gardens, universities, commercial companies, the BHL (represented by the Smithsonian Institution Library – SIL), the EDL Foundation, and other EU projects (EDIT (European Distributed Institute of Taxonomy), SYNTHESYS (Synthesis of Systematic Resources), etc.).

## 1.2 Expected results

The proposed **BHL-Europe** project will produce the following specific and measurable results:

- (1) a robust biodiversity community portal with open, distributed architecture to provide multi-language access to the digital content
- (2) ~ 20 million pages of biodiversity literature from a large number of EU Member States for display through the EUROPEANA portal
- (3) tested and validated best practice methods, standards and specifications for technology platforms, digitisation and image storage
- (4) tested and validated methodology for content enrichment from content holding institutions
- (5) tested and validated best practice workflow on implementing the BHL-Europe architecture
- (6) sustainable and persistent digital curation of biodiversity heritage literature; preservation and conservation of rare and fragile material
- (7) the integration of Taxonomic Intelligence Web tools to facilitate the search for taxa-specific biodiversity information
- (8) improved efficiency of research in the biology domain; improving access to information to non-museum biologists; building public engagement, awareness and participation
- (9) permission from publishers (not-for-profit biodiversity learned societies) to digitise previously-published content
- (10) a metadata repository and collection analysis tool for all the leading libraries involved
- (11) strategies, plans and processes for long-term preservation and sustainability of the data produced by national biodiversity digitisation programmes as part of BHL-Europe

## 1.3 List of participants

The partners for BHL-Europe include most important natural history museums and botanical gardens in Europe because they have large (digital) biodiversity libraries. All are important domain experts, and are important disseminators of the project's outcomes through their extensive participation in related community networks and their contacts with the most important target user groups. Several of the selected partner institutions combine their domain and library expertise with a strong expertise in biodiversity informatics and related IT issues making them ideal for the technological implementation of BHL-Europe (e.g. NHM, FUB-BGBM, NAT, RMCA, and NHMW). The IT expertise of BHL-Europe is completed by two commercial companies (AIT, ATOS), both of whom are highly experienced in the development, adaptation, and implementation of digital archives and libraries. These commercial companies are full partners of the network and

are responsible for most of the technical integration and standardisation of digital content. For details of the expertise see Table below and section 9 of the proposal.

It is obvious from the list provided below that the consortium includes a considerable number of partners. However, most of the actual work is done by a core group of seven partners (AP 1-7) (see also labour effort overview in section 7). The other partners are either providers of content or advisors in biodiversity and technological issues and therefore have a smaller amount of work and labour effort.

One aim of BHL-Europe is to increase the number of partners during the project to get the relevant content of biodiversity literature, to include the most important languages, and to ensure a European-wide level of awareness. Therefore, we are in contact with a number of potential partners who are interested in joining at a later stage in the project. These institutions include the Stockholm University, the Swedish Museum of Natural History, the National Herbarium of the Netherlands, the Conservatoire et Jardin botaniques de la Ville de Genève, Amsterdam University, the National Library of Latvia, and the National Natural History Museum and Botanical Garden of Portugal.

### List of Participants

Potential content providers that have no digital content at the moment but in the future are given in brackets.

AP No <sup>1</sup>	Participant name	Partic. short name	Country	Role in the project <sup>2</sup>	Expertise	Date enter project	Date exit project
1	Humboldt-Universität zu Berlin	UBER	DE	Project coordinator (Content provider) Technology provider Disseminator Domain / Language expert	Biodiversity research; project management; IT/ICT; digital libraries; large library & collections; link to community networks	M1	M36
2	Natural History Museum	NHM	UK	Content provider Technology provider Disseminator IPR expert Domain / Language expert	Biodiversity research; project management; IT/ICT; digital libraries; large library & collections; link to community networks & BHL, IPR	M1	M36
3	Narodni muzeum	NMP	CZ	Content provider Disseminator Domain / Language expert	Biodiversity research; project management; IT/ICT; PR; digital libraries; large library & collections; link to community networks	M1	M36
4	EDL Foundation	EDL	NL	Technology provider Disseminator	Connecting cultural heritage; project management; IT/ICT; digital libraries and archives; link to digital archive networks	M1	M36

<sup>1</sup> Participant number 1 is the Coordinator. The remaining participants are beneficiaries.

<sup>2</sup> The main operational role that the participant plays in the proposed project. For example: content provider, technology provider, pedagogical expert, standardisation body, evaluation, dissemination etc.

5	Angewandte Informationstechnik Forschungsgesellschaft mbH	AIT	AT	Technology provider	Project management; information management; distributed databases; IT/ICT; digital archive and OAI technologies	M1	M36
6	ATOS Origin Integration France	ATOS	FR	Technology provider	Project management; IT/ICT; information management; archive & access to digital content; OAI technologies; system integration & implementation	M1	M36
7	Freie Universität Berlin	FUB-BGBM	DE	(Content provider) Technology provider Disseminator Domain / Language expert	Biodiversity research; project management; bioinformatics; large library & collections; link to community networks	M1	M36
8	Georg-August-Universität Göttingen Stiftung Öffentlichen Rechts	UGOE	DE	Content provider Domain / Language expert	Biodiversity research; project management; digital libraries; large library; link to BHL	M1	M36
9	Naturhistorisches Museum Wien	NHMW	AT	(Content provider) Technology provider Domain expert	Biodiversity research; project management; IT/ICT; large library & collections; link to community networks	M1	M36
10	Oberoesterreichisches Landesmuseen	OOELM	AT	Content provider Domain expert	Biodiversity research; project management; IT/ICT; digital libraries; large library & collections; link to community networks	M1	M36
11	Hungarian Natural History Museum	HNHM	HU	Content provider Domain / Language expert	Biodiversity research; project management; digital libraries; large library & collections; link to community networks	M1	M36
12	Museum and Institute of Zoology, Polish Academy of Sciences	MIZPAS	PL	(Content provider) Domain / Language expert	Biodiversity research; project management; large library & collections; link to community networks	M1	M36
13	The Natural History Museum of Denmark, University of Copenhagen	SNM	DK	Content provider Domain / Language expert	Biodiversity research; project management; digital libraries; large library; link to community networks	M1	M36
14	Stichting Nationaal Natuurhistorisch Museum Naturalis	NAT	NL	Content provider Disseminator Domain / Language expert	Biodiversity research; project management; IT/ICT; digital libraries; large library & collections; link to community networks	M1	M36
15	National Botanic Garden of Belgium	NBGB	BE	Content provider Domain / Language expert	Biodiversity research; project management; digital libraries; large library & collections; link to community networks	M1	M36



16	Royal Museum for Central Africa	RMCA	BE	Content provider Technology provider IPR expert Domain / Language expert	Biodiversity research; project management; IT/ICT; digital libraries; large library & collections; link to community networks	M1	M36
17	Royal Belgian Institute of Natural Sciences	RBINS	BE	Content provider Domain / Language expert	Biodiversity research; project management; digital libraries; large library & collections; link to community networks	M1	M36
18	Bibliothèque nationale de France	BnF	FR	Content provider Domain / Language expert	Project management; IT/ICT; digital libraries; large library; link to library networks	M1	M36
19	Museum national d'histoire naturelle	MNHN	FR	Content provider Domain / Language expert	Biodiversity research; project management; digital libraries; large library & collections; link to community networks	M1	M36
20	Consejo Superior de Investigaciones Cientificas	CSIC	ES	Content provider Domain / Language expert	Biodiversity research; project management; digital libraries; large library & collections; link to community networks	M1	M36
21	Museo di Storia Naturale dell'Università degli Studi di Firenze	MSN	IT	(Content provider) Domain / Language expert	Biodiversity research; large library & collections; link to community networks	M1	M36
22	The Royal Botanic Garden Edinburgh	RBGE	UK	Content provider Domain expert	Biodiversity research; project management; digital libraries; large library & collections; link to community networks	M1	M36
23	Species 2000	Sp2000	UK	Domain expert	Project management; information management; link to community networks	M1	M36
24	John Wiley & Sons limited	Wiley	UK	(Content provider)	Project management; E-publishing; digital libraries and archives; link to publishing networks	M1	M36
25	Smithsonian Institution	SIL	USA	Content provider IPR expert Disseminator Domain expert	Biodiversity research; project management; digital libraries; large library & collections; link to community networks; leadership of BHL; IPR.	M1	M36
26	Missouri Botanical Garden	MOBOT	USA	Technology provider Domain expert	Biodiversity research; project management; IT/ICT; digital libraries; large library & collections; link to community networks; BHL technology.	M1	M36

## 2 Contribution to eContentplus programme objectives

BHL-Europe focuses on **Action 5.1. Best Practice Networks for interoperability of digital libraries** with the objective of improving the interoperability of digital libraries held by 15 natural history museums and botanic gardens, two institutional archives, and libraries across 12 EU Member States and will be progressively extendable.

BHL Europe will

- establish standards-based interoperability between digitised documentation, text, metadata and the collections, thus ensuring the interoperability of the systems for
  - enabling the use of the multilingual and cross-cultural search and retrieval of the digital content
  - facilitating user access through the common user interface of the European Digital Library (EUROPEANA)
- establish highly interoperable databases of the biodiversity-related content held by museums, archives and libraries across 12 EU Member States
- enable the actual content to be accessible and retrievable at item level by users across the European Research Area (ERA) and beyond in developing countries
- take full account of the users and their needs
- extend the range of users

The partners in this consortium have shown their capacity to contribute significantly to the achievement of these objectives and to contribute on an ongoing basis to the advancement of the European Digital Library.

BHL-Europe will bring a large amount of new material to EUROPEANA - and this material will be from a major science domain.

The collections of biodiversity literature in Europe are spread across a large number of institutions. Inevitably, there is much duplication of content. While each national government will be expected to fund scanning in their own country, it will be critical that we avoid scanning the same material in multiple locations. Duplication would waste European financial resources and deliver a very confusing set of Web sites for the user. Coordinating the scanning process at the European level will maximise the benefit of the funds for scanning, and amortise the cost of developing this collection across the whole of the community. The BHL project will be scanning materials globally and it is critical that overlap and duplication is avoided in sub-projects. Technical solutions for duplication control will be established using in particular the experiences of present large scanning initiatives at NHM, UGOE, OOELM, BnF, and SIL.

Established skills in database creation and the Web site development, currently only available in a few partner institutions (e.g. AIT, ATOS, EDL, FUB-BGBM, NAT, NHM, RMCA, UBER), will be shared across the project. This will lead to rapid development of the project, and skills development in the partner institutions both large and small.

An EU-wide project of this scale will be able to negotiate, where necessary, with rights holders at an EU level rather than on an individual national level. This will create an effective and convincing infrastructure for negotiations.

It is critical that this material – once scanned – is maintained in a sustainable way and is available in perpetuity. It is unlikely that any one country or institution will be able to manage this. Sustainability is only likely to be achieved by a consortium of major European institutions committed to work as a group to provide a sustainable infrastructure in the long-term.

In conclusion, BHL-Europe addresses the objectives of the i2010 digital library initiative of the European Commission. It represents the connecting link between the first part of the initiative (cultural heritage) and the second part of the initiative (scientific heritage). The digital material

available through BHL-Europe and the BHL (see Appendix, section 10.4.1) directly complements the digital content of the European Digital Library - EUROPEANA. BHL-Europe also addresses the objectives of the Action Plan for Biodiversity of the European Commission. The Commission considers it vital to substantially strengthen the knowledge base for conservation and sustainable use of biodiversity, which requires strengthening the European Research Area (ERA) and communication and interoperability of data (COM (2006) 216 final). BHL-Europe aims to make research data on biodiversity interoperable and accessible to all interested parties. The interoperability and accessibility of existing environmental data and information is also a challenge addressed by the Shared Environmental Information System (SEIS). SEIS is a collaborative initiative of the EC and the European Environment Agency to establish an EU-wide environmental information system.

### **3 European dimension**

At the core of the BHL-Europe project is the relatively simple premise of making large amounts of scanned biodiversity literature freely available over the Web and indexed with semantic tools. The outlets for this material will include national portals, the BHL portal and, most importantly, EUROPEANA. The process of making significant scientific content 'more accessible and usable and exploitable' by its addition to EUROPEANA will transform EUROPEANA from an arts and humanities portal into a broad European cultural resource incorporating a currently under-represented scientific dimension.

In particular, the addition of biodiversity materials in many European languages will point to the significance of European science over the centuries. The 18th and 19th century tradition of scientific investigation and expeditions, together with our colonial history, means that our literature collections contain some of the most significant and important material in the world. These collections contain vast amounts of material on the biology of the developing world, and making this material freely available on the Web will enable people in the developing world to gain access to this knowledge for the first time. This will help the EU respond to its commitments to the Convention on Biological Diversity (CBD) – a key aspect of the CBD is the exchange of information 'that shall include exchange of research results and repatriation of information' to the developing world.

EUROPEANA has already established a reputation as being multicultural and multilingual. BHL-Europe will help to broaden the concept of 'culture'. The multilingual nature of much of this literature will also enable the development of different approaches to issues related to language. An improved understanding of multi-language documents and multi-language OCR techniques will result from work in this project, and will be available to all users of digital objects.

BHL-Europe will adopt and use established standards wherever possible - produced by EUROPEANA, the BHL and others - to ensure maximum interoperability of biodiversity literature with other systems and services.

### **4 Content**

The content providers within the BHL-Europe consortium will create a critical mass of digital content of high quality representing the biodiversity domain. The content providers have been selected on the basis of their ability to contribute relevant biodiversity and taxonomic literature (zoology, botany, palaeontology).

A key selection criterion is that the content available to users should not be restricted by proprietary third-party rights or any other constraints, which would limit its use in an open access environment using Creative Commons licences. The digital content must either be out of copyright and in the public domain (e.g. Germany will provide everything published before 1910) or else the content

contributors must have permission from IP owners (for instance, not-for-profit and institutional publishers) to give open access.

The initial selection of public domain material is not a limitation of the project because systematic biology depends more than any other natural science upon historic literature, as mentioned above (section 1.1.1). Another reason to focus on historical literature is that many old and important monographs are themselves inherently very rare, fragile or in need of conservation. This makes “hands on” access very difficult. This project will substantially reduce the need for handling of these rare and valuable materials. Detailed selection of content and work on the selection criteria is also part of WP2 of the project and the methodology is discussed below (section 7 Project Work Plan).

In the following table, the digital content identified as available reflects the situation as it exists in May 2008. As many consortium partners are just beginning the digitisation of their biodiversity literature, the amount of digital content will increase significantly during the project lifetime. Therefore, an estimate of the number of pages available at the end of the project is given in brackets. This indicates the rate of development of European biodiversity scanning initiatives. Some countries (such as Germany) have not yet started digitisation, but are preparing to start during the first year of the project and will contribute a significant quantity of digital content.

In addition to the increase in the amount of content from each content provider during the project’s lifetime, we expect new partners to provide another source of content. These new partners may be other important libraries in countries not yet engaged in the project or learned societies. Following the establishment of the harvesting procedures and the development of the Memorandum of Understanding, this will be an important task of WP2 (section 7 Project Work Plan). Attraction of new partners and the initiation of national scanning projects will be particularly important in mobilising the very significant amount of biodiversity literature in Spanish, Italian and Portuguese that currently is scarcely available on the Internet (see table below).

Quantity and Quality of the Content								
Provider <sup>3</sup> / Location	Type <sup>4</sup>	Quantity & Definition <sup>5</sup>	Format & Quality <sup>6</sup>	IPR <sup>7</sup>	Current Use <sup>8</sup>	Existing Metadata <sup>9</sup>	Language	Additional comments
UBER / Berlin, Germany	Page images, metadata	(~800,000 pages)	TIFF, JPEG, OCR; colour, 600 dpi	Creative Commons/ Science Commons, or public domain	Not applicable	Not applicable	German	
NHM / London, UK	Page images, metadata	~400,000 pages	TIFF, JPEG, OCR; colour, 600 dpi	Creative Commons/ Science Commons, or public domain	Mainly scientists; historians and artists; 13,000 visitors per year	Taxonomic metadata; bibliographic metadata	English	
NMP / Prague, Czech Republic	Page images, metadata	1000 pages (20,000 pages)	TIFF, JPEG; colour, min. 300 dpi	Public access but restrictions to commercial use	Mainly Palaeobotanists (~150)	Unstructured	English	
FUB-BGBM / Berlin, Germany	Page images, metadata	(~700,000 pages)	TIFF, JPEG, OCR; colour, 600 dpi	Creative Commons/ Science Commons, or public domain	Not applicable	Not applicable	German	
UGOE / Göttingen, Germany	Page images, metadata	160,000 pages (450,000 pages)	TIFF, JPEG, PDF; colour and greyscale	Public domain or copyright GDZ	Mainly scientists	Taxonomic metadata; bibliographic metadata	Various languages	
OOELM / Linz, Austria	Page images, text, metadata	~150,000 pages (800,000 pages)	PDF, OCR; colour, greyscale	agreement with the cooperating institutions to provide online access to content	Mainly scientists	Own system	German	
HNHM / Budapest, Hungary	Page images, metadata	~20,000 pages (~50,000)	TIFF, PDF, OCR; bw – 600 dpi, greyscale	Free access for scientific and education purposes	Mainly scientists	Bibliographic metadata	English, German, Italian, French, Latin	

<sup>3</sup> Short name of the participant who provides the content

<sup>4</sup> E.g. Text, image, movie, sound, music etc.

<sup>5</sup> E.g. 1,000 film clips, 2 million pages, 20,000 books etc.

<sup>6</sup> E.g. Format - JPEG, MPEG, QuickTime, HTML, PDF etc., Quality – Resolution, sampling rate, colour/greyscale etc.

<sup>7</sup> Access rights to use the content in the proposed project e.g. public domain or license

<sup>8</sup> Describe current users of the content, including the number (e.g. registered users)

<sup>9</sup> Describe fields, languages and structure of the metadata

		pages)	and colour – 300 dpi					
SNM / Copenhagen, Denmark	Page images, metadata	~5,000 pages (~100,000 pages)	TIFF, JPEG; 800 x 534 (pictures), 800 x 225 (text)	Public domain	Mainly scientists	Bibliographic and taxonomic metadata	Danish, English, German, French	
NAT / Leiden, The Netherlands	Page images, text, metadata	100,000 pages (200,000 pages)	JPEG2000, PDF, OCR-PDF; up to 600 dpi	Most content owned by NAT; no restrictions for private and educational purposes	Popular science and education: 3,000,000 users per year; amateur scientists and scientists: 700,000 users per year	extended Dublin core; taxonomic metadata; bibliographic metadata	English, Dutch	
NBGB / Meise, Belgium	Page images, text metadata	1,500 pages (~50,000 pages)	TIFF, JPEG, OCR; colour, 600 dpi	Creative Commons/ Science Commons, or public domain	Internal use, stakeholders of the wider public	Not applicable	French, Dutch, English	
RMCA / Tervuren, Belgium	Page images, text, metadata	~100,000 pages	TIFF, OCR, PDF, PNG, JPG	Use regularly Creative Common for other project, property of the institution but as public institution considered as public domain	External and internal use by scientists and other stakeholders of the wider public	Mainly scientists	French, English, Dutch, German, Danish, African local languages	We also have the archives of large sampling expeditions in Africa
RBINS / Brussels, Belgium	Page images, text, metadata	~100,000 pages	TIFF, OCR, PDF		Scientists, internal use	Bibliographic and technical metadata	French, English, Dutch	
BnF / Paris, France	Page images, text	>1,500,000 pages (~2,000,000 pages)	TIFF, JPEG; 300 dpi	Public domain	200,000 visitors per month; researchers and general public	XML / Dublin Core	French, English, Latin, German, Others	
MNHN / Paris, France	Page images, metadata	100,000 pages	TIFF, JPEG; colour, greyscale, 300 dpi	Public domain	2,500 users since March 2008 (portal opening)	Dublin core	French	
CSIC / Madrid, Spain	Text, images	2,500 pages	TIFF; PDF min 300 dpi with OCR, greyscale and colour when necessary	Public access but restriction to commercial use	Mainly scientists but also amateur scientists	Bibliographic metadata	Spanish, English	
Sp2000 / Reading, UK	Database: Catalogue of Life, Annual	1,100,000 species,	MySQL/PHP Database	License	Indexing Biodiversity Data,	Not applicable	Multiple	Used for Taxonomic

	Checklist	increasing each year, ISSN 1473-009X			40 national portals, 10 M hits per month			Intelligent Tools
RBGE / Edinburgh, UK	Page images, text	15,000 pages (~100,000 pages)	High quality (details to be confirmed)	Freely available but RBGE retaining copyright	Not applicable	Not applicable	English	
SIL / Washington, USA	Page images, metadata	~3,600,000 pages (15,000,000 pages)	JPEG, OCR; colour, 600 dpi	Creative Commons/ Science Commons, or public domain	Mainly scientists, 13,000 visitors per month	Taxonomic metadata; bibliographic metadata	Mainly English	
<b>TOTAL</b>		<b>6,255,000 pages (20,972,500 pages)</b>						

#### 4.1 IPR issues

The consortium has the necessary licensing and clearing arrangements in place for the Intellectual Property Rights (IPR) arising from the proposed project and this will ensure the far wider use and dissemination of the project output across all 27 EU Member states.

Participant 2 (NHM) has extensive experience of managing IP issues in a range of EU and other projects – for example SYNTHESYS, EDIT, BHL, etc. Participant 2 has a dedicated IP officer, who will work on this project, and will also provide IP support to WP2 of EUROPEANA.

BHL-Europe will digitise material that comes under two main groupings/categories:

**Public domain** – this is where the IPR on the material has expired. Conditions vary from country to country within the EU and we will be seeking advice to ensure that material is definitely out of copyright in the country of publication. This will ensure that any public domain materials will remain public domain and available on an open access basis. Public domain material can be reused or exploited by anyone who wishes to use it including educational, non-commercial, and commercial uses.

**Rights held by other organisations** – this is where the IPR for material is still owned by an organisation or individual. The English language BHL project has already been successful in persuading not-for-profit organisations, learned societies, and institutional publishers to allow 49 different journals and series, some containing many thousands of pages to be made available through the BHL. This material will be available on an open access basis. Many more titles are currently under negotiation. Rights agreements reached by BHL-Europe will be for global open access – thus including BHL and EUROPEANA – and will be managed through Creative Commons (CC) licences. The Rights Holder will retain some rights in the material – as managed by the CC license agreed – but this will only prevent commercial reuse in some cases.

We have pro forma documents and agreements available already, and would be able to start making further agreements with rights holders as soon as the project starts.

#### 4.2 Multilingual and/or multicultural aspects

##### Background

The biodiversity literature is available in many languages – overwhelmingly European in origin – and must now be made accessible in many more languages. Local users of the literature will, wherever possible, use the EUROPEANA or BHL Portal in their own language in order to gain full benefit to the literature available. Differences in language, geography, population, and social and economic conditions are significant factors affecting the ability of users to exploit the BHL resource. European skills and experience in managing multi-language and multilingual resources gives BHL-Europe the capability of making this material available as widely as possible.

##### Multilingual Web sites

EUROPEANA will provide a multi-language interface, as will the BHL portal. However, the ability to exploit documents in multiple languages will need the design and development of excellent portal tools and improved OCR systems. The most obvious, but short-term, solution would be to translate the Portal interface into other languages. Portal languages anticipated by the end of the project are Czech, Danish, Dutch, English, French, German, Hungarian, Italian, Polish, Portuguese, and Spanish. However, a much deeper level of language integration is necessary, such that a user will have the facility to search for a biological species in either Latin or their local language (i.e. Dutch) and return a page originally written in Polish.

Portals and Web sites will offer services via interfaces that support multiple languages. OCR processing of scanned images will identify key words, names, relevant terms, etc. and these will be



stored as metadata. Therefore, the search and other utilities will be multi-language and results will be returned in the original language of the document. It is not the intention to translate original texts into multiple languages at this stage, but this may be technically possible in the medium-to-long-term.

### **Multilingual indexes**

In biodiversity, the most important aspects of meta-information are the names of organisms, groups, and named locations (the latter uses geographical, bio-geographical, or ecological terminology). Unfortunately, these areas are not stable – index terms may have different meanings depending on who applied them and when (e.g. Germany, Liliaceae). Moreover, ecological and taxonomic class names represent scientific concepts, thus parallel partly-overlapping hierarchies may exist between different semantic and ontological sets. BHL has constructed a ‘Taxonomic Intelligence (TI)’ tool (a form of thesaurus) to form a common source for both indexing of resources and for the Portal's user interface (e.g. a convenient taxonomic browser instead of free text fields). The thesaurus has to be powerful enough to treat various semantic relations such as synonyms and hierarchies. Currently, the thesaurus has a preponderance of Latin binomial scientific and some vernacular names, mainly English. It is critical that a true multilingual version should now be developed. BHL-Europe, through its many language partners, will be able to add many more local non-English names to the TI database. AIT and ATOS, in particular, will support these language partners with their proven expertise in the implementation of multilingual thesauri. Eventually, it will for the first time enable the corpus of biodiversity material to be searched in any European language using local names for species and locations.

### **Multi-language optical character recognition**

Currently the parsers for optical character recognition (OCR) work in a single language. It is possible to load two languages into a parser – for example, English and Latin – but this becomes unmanageable with the large number of languages used within the EU. We will have to ensure that the metadata for each document contains appropriate language markers and that separate parsers are used to OCR each document. This approach is not in wide use at present, and we will work with partners in related networks and projects to investigate how to optimise this approach to OCR. For example, we will have the opportunity to test and use a variety of appliances and services developed by the EU-funded IMPACT project (Improving Access to Text), as part of their outreach activities from 2010 onwards. Effective use of OCR will enable ‘deep’ access to documents through the actual text of the original. Successfully implemented, this technology will be highly valuable to other EDLNet projects which encompass digitisation and OCR of large quantities of material.

## **5 Impact**

### **5.1 Analysis of demand**

80 participants from 22 countries (including 7 EU countries) met in London in February 2005 to discuss the status and future of access to the taxonomic literature and what needs to be done to provide the widest possible access to the user communities. The participants were taxonomists, librarians, publishers, representatives of learned and professional societies, private foundations and government agencies, and specialists in information and communications technology. The lack of access to the published literature of biodiversity was identified as one of the principal obstacles to efficient and productive research, outreach, and education. The largest part of biodiversity knowledge is held in a few libraries, and is only available to the few scientists, who can gain direct access to these collections. This biodiversity literature is effectively unavailable for wide use by a broad range of potential users, including those in research, education, taxonomic study, biodiversity conservation, protected area management, disease control, and maintenance of diverse ecosystems services.

Currently the study of these distributed collections is difficult, time-consuming, and expensive. Requests for direct help and support in obtaining access to important literature are frequently received by partner libraries, which do not have the facilities to meet them. Researchers and students from developing countries are particularly disadvantaged, yet in many ways their need is greater since both biodiversity, and the threats to it are much higher in developing countries than anywhere else.

Recent analysis of demand (Deliverable 5.18 of EU-funded EDIT project; a survey based on a questionnaire filled by 64 scientists from 19 countries, 10 of which were within the EU) indicates that the majority of biodiversity and biological scientists need a multilingual interface to the literature, with sophisticated search and filtering functionality providing an online repository of references relevant to taxonomic research together with the original text content of articles. Users clearly want to have access to the literature itself (see also Appendix, section 10.4.3). As species names are an integral part of the biodiversity information in the literature, search, and filter functions of the interface should include name-finding tools. At the moment, other European projects and networks established in the biodiversity community focus on species names (GBIF, Species 2000), data standards for biodiversity information (GBIF, TDWG) or catalogues of content (EDIT – Virtual Library of Taxonomic Literature, see Appendix, section 10.4.2).

In addition to these large-scale projects, smaller digitisation projects are present in a number of EU member states digitising the actual literature (see content analysis above). Users of this digitised material are appreciative of the service, as demonstrated by the tenfold increase of visitors to related Web pages and portals (e.g. Naturalis, AnimalBase). The BHL portal at [www.biodiversitylibrary.org](http://www.biodiversitylibrary.org) had over 13,000 unique visitors per month during the first months following the release of the portal. The BnF report of 200,000 visitors per month for the Gallica portal ([gallica.bnf.fr/](http://gallica.bnf.fr/)). However, to date there is no coordinated effort in Europe which combines the above approaches while providing full-text access to taxonomic literature via a multilingual Web interface.

## 5.2 Target users and their needs

The target user groups who will be able to exploit open access to the biodiversity literature will be:

- A. **Scientists** – working in the field of biodiversity, but also in many related biology fields including environmental studies, ecology, environmental policy, forestry, water management, epidemiology, plant and animal diseases, and horticulture. These scientists will be at varying levels of training and professional development, from school to PhD and beyond into research.
- B. **The Developing World** – we expect access to this material to provide substantial benefit to scientists, teachers, policymakers and other groups in developing countries. This material has never before been readily available to developing countries, and will have a huge impact on capacity building and development.
- C. **Education** – The material available through BHL-Europe is an important but inexpensive resource for educational purpose. The biodiversity literature, freely available on the Internet, provides original source material for teachers and students and complements the content of textbooks and other teaching resources.
- D. **Environmental / Conservation agencies and Governmental officials** – In addition to purely taxonomic information, biodiversity literature from previous centuries highlights recent changes in ecosystems and the environmental deterioration caused by human interventions. Thus, this material is expected to be an important resource in building the Shared Environmental Information System.

- E. **Arts, humanities and social science** – these materials are able to shine a light on the sociology and history of science. In addition, many of the older texts have exquisite artistic representations of plants and animals currently little-known outside the scientific world.
- F. **The general public** – the public is interested in a wide range of biological and environmental issues. The material in these collections has never before been available to the public on a large scale, and we expect this to be a substantial group by the end of the project.

As many consortium member institutions are large natural history museums, the target users are represented in the form of staff members, scientists, students, and members of the public who visit the exhibitions. Consortium members are experienced in tracking, communicating and surveying their museum users, and expect to easily attract volunteers for tests and validation of the project results. Similarly, the involvement of the other target users (developing world scientists, policy makers, etc.) through the work of the large natural history museums should be relatively easy. Participant 1 and 2, for example, both provides artists the opportunity to work in their collections. Most museums also have an archive service dedicated to historic documents, which brings historians and social scientists to these museums.

Target user description	Needs	Feedback methods	Country coverage
A. Scientists	Taxonomic descriptions of species; biodiversity data of specific regions in the last centuries; full-text searching; taxonomic intelligence	Consortium, review group, online questionnaire, direct feedback	Consortium member states, ERA, Global
B. Developing World	Taxonomic descriptions of species; biodiversity data of specific regions; full-text searching; taxonomic intelligence; resource for teaching materials	Online questionnaire	Global
C. Students of different levels (primary to academic)	Reliable and meaningful information and relevant images on biodiversity; minimal time to aggregate information from different sources; research resource	Review group, online questionnaire, direct feedback	Consortium member states
C. School teachers	Resource for teaching materials	Online questionnaire, direct feedback	Consortium member states
D. Environmental and Conservation agencies / Government officials	Information on impact of climate change, environmental deterioration and human interventions	Online questionnaire	Consortium member states, ERA
E. Artists	High quality images of animals and plants	Online questionnaire	Consortium member states, global
E. Historians	Historical information on science and scientists	Online questionnaire	Consortium member states, global
F. General public	Direct online access to comprehensive information not currently publicly accessible to help raise the awareness and appreciation of the biodiversity heritage	Online questionnaire, direct feedback	Global

### 5.3 Critical Mass

BHL-Europe will include a variety of content providers representing different institutions, countries, and languages. They are ready to provide digital content (see Table in section 4) subject to a wide range of workflows and processing techniques. This helps to draw valid conclusions from the implementations planned during the proposed project. BHL-Europe is poised to find best

practice guidelines for data standards, semantic enrichment, multi-language issues, and integration of digital content in BHL-Europe. Once the system is working and ingestion procedures are established, further increase of digital content is significantly facilitated. Eventually, the large-scale implementation of BHL-Europe will lead to a very high quantity of digital content of biodiversity literature.

#### 5.4 Added Value

BHL-Europe's added value lies in its integrative role. There are currently a large number of small projects digitising biodiversity material in numerous institutions across the EU. These projects are not using common standards or interfaces and are currently not interoperable. This is extremely confusing for the user and is a very inefficient use of resources, since it is almost certainly the case that the same text is being scanned in a number of different centres e.g. early works from Linnaeus can be found on a number of Web sites via a simple Google search.

BHL-Europe will provide integration and interoperability by:

- establishing, and communicating common standards for the digitisation of biodiversity materials in Europe
- making that material available through at least two portals – EUROPEANA and the BHL – and thereby simplifying the situation for end-users
- developing value added tools – taxonomic intelligence, multilingual interfaces, improved OCR, etc. – and making these tools available to the community
- supporting European institutions in their national bids for funding for digitisation of biodiversity materials
- linking with the global BHL project to ensure that these biodiversity materials are available globally, and availability is sustainable in the long-term

#### 5.5 Success indicators

Indicators	Expected Progress		
	Year 1	Year 2	Year 3
Number of accessible pages of biodiversity literature	10,000,000	15,000,000	20,000,000
Number of content providers	30	35	40
Number of portal languages	2 <sup>1</sup>	6 <sup>2</sup>	11 <sup>3</sup>
Hits on Web site (Page-views)	500,000	2,000,000	5,000,000
Case studies of successful usage of the material by non-scientists	5	10	15

<sup>1</sup> English, German; <sup>2</sup> + French, Italian, Portuguese, Spanish; <sup>3</sup> + Danish, Dutch, Czech, Hungarian, Polish

#### 5.6 Dissemination and awareness

##### 5.6.1 Dissemination and demonstration

BHL-Europe will produce a consistent set of communication activities to ensure that project results become known and used by the target users identified above. The experienced team from the National Museum in Prague will be responsible for leading the dissemination activities. Eight people from that partner with various qualifications ranging from biodiversity and IT experts to PR specialists will be involved. Communication will take place at three different levels to ensure dissemination for awareness, dissemination for understanding and dissemination for action. At the beginning of the project, a dissemination plan will be developed, which will define all dissemination activities in detail. This plan will be updated regularly, to incorporate new opportunities to disseminate BHL-Europe project results. Reaching the target groups of the project

and stakeholders, we foresee various mechanisms for dissemination and awareness raising activities:

- dissemination of the project results through BHL-Europe members (using already existing networks of European scientific organisations included in the BHL-Europe consortium and within the considerable number of professional and special interest organisations where they are members, see 6.1)
- dissemination through BHL-Europe Web site and other promotional materials (fact sheets, flyers, self-running demonstrators, posters, newsletter, etc.)
- dissemination through papers in professional journals, short articles in newspapers, texts in various online encyclopaedias and information services (Wikipedia) and blogs
- dissemination through presentations and demonstrations at conferences and other relevant events

The dissemination concept will work towards enhancing the image and awareness of BHL-Europe, its products, and services using structured campaigns to reach the target audiences. The expertise will include: corporate communications, media relations, product publicity, event management, exhibitions, print production and Web site development, workshops, and training.

We will develop the following major items:

- Web site design for internal and external communication – implementation of this design will be in cooperation with WP3
- design and produce BHL-Europe project promotion kit (flyers, folders, posters, presentations)
- design and development of a newsletter and mailing list

The BHL-Europe promotion kit will be used at conferences and public events to present BHL-Europe. An online database will be created to identify relevant events and to identify the relevant consortium member to attend the event and present BHL-Europe and its result. Workshops will be offered for potential new partners to explain the workflow of BHL-Europe and to illustrate the functions of the portal. At the moment, the following types of conferences and events are identified as relevant for BHL-Europe to present the results:

- annual meetings of the large national and European scientific societies / learned societies (e.g. Systematics Association, Linnean Society, Federation of European Societies of Plant Biology, European Society for Evolutionary Biology, Deutsche Zoologische Gesellschaft)
- meetings and conferences of important community based projects (e.g. GBIF, EDIT, SYNTHESYS, CoL/Sp2000)
- meetings and conferences of environmental organisations (e.g. BUND, WWF)
- international conferences on environmental and biodiversity issues (e.g. COP9)
- meetings and conferences on e-Learning (e.g. ICVL – International Conference on Virtual Learning), ICT (e.g. IADIS - International Association for Development of the Information Society), and related topics
- public events on biodiversity and science (e.g. events associated with national and international biodiversity days, science nights, Darwin 200 celebrations, etc.)

In addition to that, we will also identify institutions for educational development and teacher training (e.g. the State Institutes for School and Media in Germany) and establish cooperation with BHL-Europe. This will enable the involvement in training courses for teachers and contact to schools and universities to introduce the BHL-Europe results to this target user group.

### **5.6.2 Exploitation plan, activities & sustainability**

WP5 will focus on communicating the existence of this new biodiversity content to a range of potential users. EUROPEANA and the natural science networking projects (EDIT and SYNTHESYS) will give us access to a wide number of potential users. However, it will be a challenge to reach the general public and education communities and we will be developing a detailed exploitation plan to reach all these user communities.

One of the tasks of the BHL-Europe Communications Working Group (see section 8.6) will be to establish a communications strategy, so as to assure that the project will reach all its target users. In addition to contributing to the communications plan, each project partner will be expected to exploit the content and results locally, and help identifying related networks or organisations that might be interested in using BHL-Europe content or services.

In the nature of open access cultural repositories and portals (such as EUROPEANA and BHL), it is difficult to develop a sustainable business model which is dependent on income from the content and services created. It is the very nature of this content model that opportunities for exploitation will lie with the users, not the creators of the service. We will look closely at the options for a sustainable business model. Possible solutions might include:

- create a Foundation (like the EUROPEANA model)
- establish an EEIG – European Economic Interest Grouping. The purpose of the grouping is to facilitate or develop the economic activities of its members by a pooling of resources, activities, or skills. This will produce better results than the members acting alone
- obtain commitments from natural history institutions to pay a ‘fee’ or subscription to sustain the service for their scientists and others
- seek committed funding from national governments or the EU (like facilities such as EMBL, ESA, etc.)
- establish an endowment and use the income to maintain access

There is a strong dependency with WP3 to ensure that these activities are aligned with the long-term storage and service concept for BHL-Europe. Following Month 24, work in WP2 on service costs and WP3 on technical costs will lead to a cost model and enable WP5 to begin to assess the options for sustainable funding.

### **5.6.3 Evaluation plan & activities**

We will put in place Web-based evaluation tools to survey the users of the portal – determination of the users, country of the users, most interesting content etc. Web 2.0 applications, particularly wikis, blogs, and podcasts will be adopted by BHL-Europe service.

We will develop online questionnaires to identify user requirements, preferences, experiences, benefits, and unmet needs. User evaluation will be carried out twice, the first time at the end of Month 12 to enable the analysis of demand and service elements of the project. This will be fed into WP3 to ensure that we focus on those key components that support user needs. The second evaluation in Month 24 will test and validate the impact of the project. It will identify key features which are highly valued by users and will feed into the final revised set of best practice guidelines and components, implemented by WP3 and published by WP2.

## **6 Networking**

### **6.1 Networking Capacity**

At present, the BHL-Europe consortium brings together 26 institutions from 12 EU Member States, and two institutions from USA. The most important European institutions for the scientific community and the public interested in biodiversity (i.e. large national history museums) are

included from the beginning in the BHL-Europe consortium. This significantly helps to raise awareness and promotes the uptake of the projects results from the majority of the 27 EU Member States.

Major players in other domains and specific digital library networks or projects are already included in the BHL-Europe consortium. These networks or projects include EUROPEANA (EDLNet – European Digital Library Network), CETAF (Consortium of European Taxonomic Facilities), GBIF (Global Biodiversity Information Facility), TDWG (Biodiversity Information Standards), ENBI (European Network for Biodiversity Information), EDIT (European Distributed Institute of Taxonomy), SYNTHESYS (Synthesis of Systematic Resources), Catalogue of Life (Species 2000), STERNA (Semantic Web-based Thematic European Reference Network Application), ENRICH (European Networking Resources and Information Concerning Cultural Heritage), BHL (Biodiversity Heritage Library), and EOL (Encyclopedia of Life). In addition to this existing integration of the proposed project, BHL-Europe is already in contact with other important projects like Key2Nature and IMPACT (Improving Access to Text). The further development of BHL-Europe's networking capacity is driven by the extensive network of contacts brought by the project partners. Some partners have accumulated large contact lists during their involvement in national and international projects, and integration with existing networks and activation of existing contacts will guarantee BHL-Europe an effective dissemination of project results.

BHL-Europe builds on biodiversity literature and should be a comprehensive resource for biodiversity information. Thus, the increase of content and the attraction of further content providers is an essential component. There is an important task in WP2 dealing with that issue (see section 7). Based on the analysis of content, gaps in the record of biodiversity literature corpus will be identified. Using the network and contacts mentioned above, institutions and organisations which hold the missing corpus of literature will be identified. WP2 of BHL-Europe will assist these potential partners in the sourcing of funds and implementation of their scanning operations to enable their contribution of content to the digital repository of BHL-Europe. At the moment, we are in contact with several institutions that are interested in contributing to BHL-Europe at a later stage in the project. These institutions include the Stockholm University, the Swedish Museum of Natural History, the National Herbarium of the Netherlands, the Conservatoire et Jardin botaniques de la Ville de Genève, Amsterdam University, the National Library of Latvia, and the National Natural History Museum and Botanical Garden of Portugal.

## 6.2 Clustering Activities

The BHL-Europe project has a strong interest in a number of the broader agenda items being addressed by the Commission and can contribute to clustering activities through:

- supporting the work of European scientists by making resources and expertise more readily available through the Web. This could be achieved by providing support for training of taxonomists in the EDIT and SYNTHESYS projects.
- establishing standards for Web access to biodiversity and related cultural materials. These will be disseminated through EUROPEANA and other digital library programmes at a European and national level.
- providing expertise in developing models for sustainability of important datasets; hosting critical scientific datasets at the European level. This will involve working with large-scale projects which will need to store substantial quantities of data for the long-term e.g. LifeWatch, ESA, etc.
- providing experts to assist with capacity building and information sharing with the developing world. BHL-Europe will link with key institutions in developing countries to optimise and sustain this process, and also with individual government-supported capacity building activities.

- exploiting European skills in languages to improve the communication of Web sites with non-English speakers. BHL-Europe will engage with developing transnational clusters of language groups, where our experience is relevant.

## **7 Project work plan**

### **7.1 Introduction and general description**

The project life cycle encompasses four phases, which are described below in more detail.

#### **Phase 1: Analysis of status quo (Months 1-3)**

Initial work will consist of evaluation of the technological solutions available, and assessment of the requirements of each content provider. We need to review the state-of-the-art technologies used by EUROPEANA and BHL for processing digital content, and we will include some of our experienced technology partners (e.g. NHM, AIT, ATOS, NAT, RMCA) in that process. Simultaneously, each content provider will provide their specific requirements: what do they expect from BHL-Europe; how the digital content should be handled (technically, scientifically, and legally); what data standards and specifications they use. This information will fill out the data provided already in section 4 of this proposal. This technology and content-related information will be analysed in order to get a comprehensive picture of the variety of implementation approaches, their advantages, and their disadvantages.

In the first phase we also need to develop the prerequisites for the management of the content, i.e. the content analysis tools and bibliographic databases. These databases will show the current position of the national scanning initiatives and, by merging the metadata from each partner, create a virtual taxonomic library database.

While WP2, WP3, and WP4 are analysing the state-of-the-art technologies and content holder requirements, WP5 will be working on the communication plan. As a result of the first phase, the project's Web site will be published and the first promotion and presentation kits will be developed. Furthermore, a Communications Working Group will be established, responsible for revising dissemination strategies and deciding on activities related to communication of project's outcomes.

#### **Phase 2: Consensus building (Months 4-6)**

In the second phase, the most appropriate implementation approaches identified in the previous analysis, will be discussed in the consortium. During discussions with each partner we will establish: a consensus of best practice guidelines; how to process biodiversity literature metadata; and how to ingest the metadata and files into the portals. Ultimately, we will have identified the best practice workflow and data model, and we will agree the Memorandum of Understanding (MoU) and IPR working documents – our first important milestone. Consensus building is also important for the communication activities. Therefore, a plan will be developed during this phase and consortium partners will agree the communication plan at the end of phase 2.

#### **Phase 3: Enrich content and test of implementation approach (Months 7-24)**

In Phase 3 with our best practice guidelines in place, we can start harvesting content from content providers. We will build a prototype system, based on the data from the consortium partners who have large amounts of German language content available now (e.g. UGOE, OOELM, HNHM, SNM, BnF, and SIL). This will test the implementation approach is working for metadata and digital content which has been subjected to a wide range of workflows and processing techniques. This will be the first test of the model in a real-life context. Following the development of the German language prototype, the key components (incl. taxonomic intelligence) will be developed and released at the end of the second year of the project – an important deliverable of BHL-Europe.



Also in this phase, BHL-Europe will start content analysis utilising the databases and tools developed in Phase 2, and begin support for the management of the scanning initiatives of each partner. It is important to eliminate duplication during the scanning process. The bibliographic database will contain information on monographs and serials that have been scanned in the past and are available in the format defined in the MoU, which are to be included in EUROPEANA and BHL-Europe. In addition, information will be provided about biodiversity literature in the process of being digitised and identifying who is the partner responsible for scanning that corpus. Eventually, the database will contain information on all relevant literature that needs to be scanned and will identify the partner who will be responsible for providing this literature in the future. If there is no BHL-Europe partner with some of the critical content, appropriate content holders will be identified and encouraged to join the consortium to provide this content. This system ensures that every content provider and even potential content providers can check before they start digitisation to see if the material is already in the queue. The local digitisation processes can be planned accordingly and duplication is reduced to a minimum, ensuring the effective use of the local resources available in each partner institution.

The partner network will need to be extended to fill gaps in the biodiversity literature corpus as mentioned above. In addition, learned societies and commercial publishers that are interested in providing their content or parts of their content through EUROPEANA and BHL-Europe need to be identified. This process will start after we have a first overview of content already present in each partner library and the content required but not present in any of the consortium libraries (gap analysis). The content analysis and content management processes will continue through Phase 4 of the project.

#### **Phase 4: Evaluation, validation, exploitation, and demonstration (Months 25-36)**

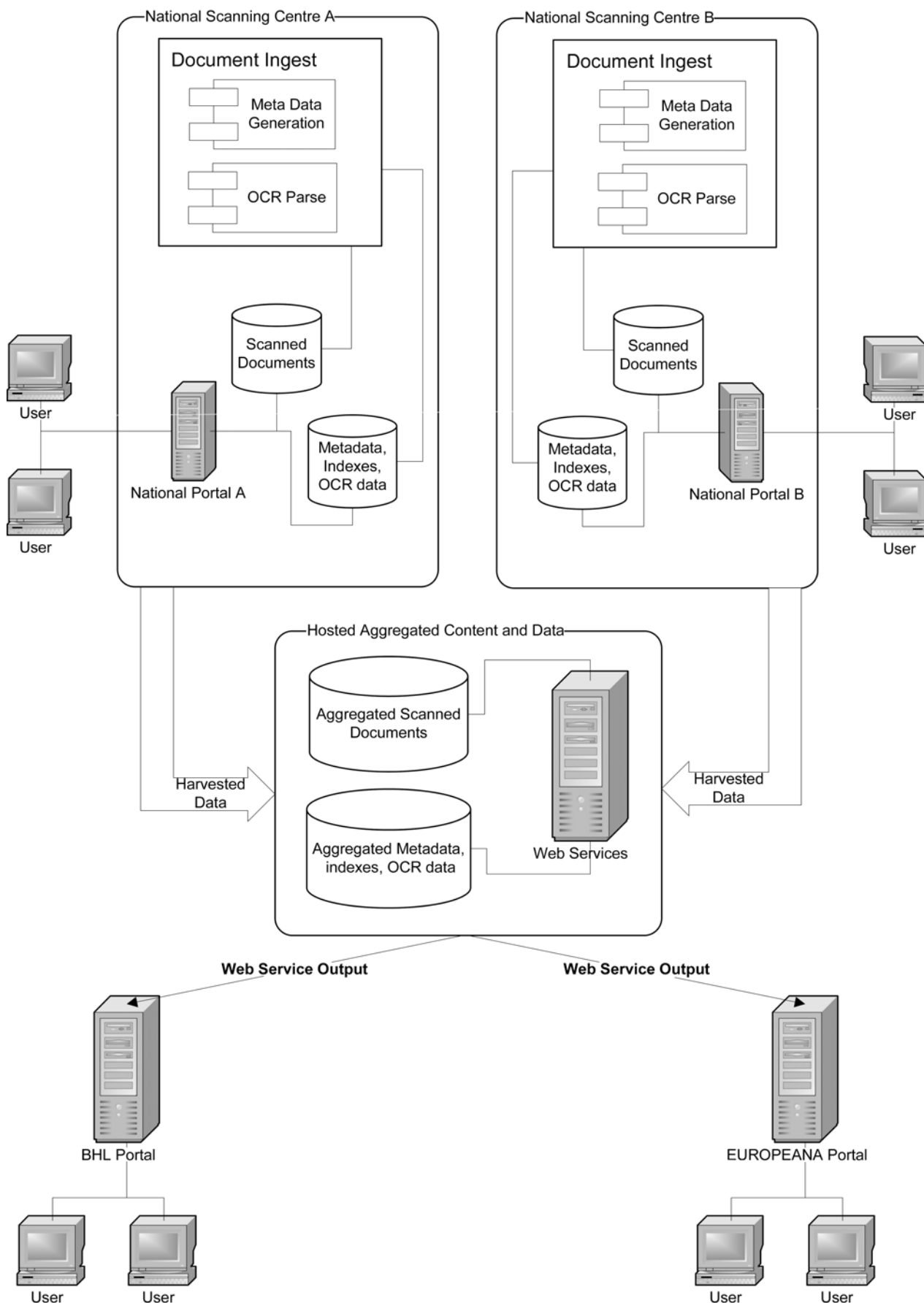
In addition to continuing Phase 3, Phase 4 mainly focuses on evaluation, validation, exploitation, and demonstration of BHL-Europe. This work is based on preliminary work done in the first phases of the project with the establishment of the communications strategy, and the development of online evaluation questionnaires. After the release of the key components, the online questionnaires need to be completed by a number of users from all target user groups of BHL-Europe. The evaluation of the questionnaires will help to validate the best practice agreed, and show where to improve the outputs of the BHL-Europe partners. The results of the evaluation process will be fed into the development of the EUROPEANA and BHL portals. Having reviewed the practical implementation, we will publish our tested and validated best practice approach for use in other projects. Large-scale demonstration activities at important conferences and public events, as well as workshops for particular target users, will increase the community and public awareness of BHL-Europe. Based on the sustainability strategy, technical solutions will be developed to ensure long-term sustainability and accessibility of the metadata and content beyond the end of the project – in perpetuity.

#### **Technical/operational framework**

Each nation or language group may have one or more scanning centres where original material will be scanned and then processed. OCR and various tools will extract and match relevant data in order to build a set of metadata that will allow for the development of multilingual and taxonomic intelligence tools within various portals.

The data will then be harvested and stored in an aggregated data centre, which provides resilience for the service and copies of all data for disaster recovery purposes. In addition to the aggregated data, Web services will be utilised or developed to allow access to this data for inclusion in various portals including EUROPEANA and BHL. This model would allow for any future, and as yet undefined, portals or Web sites to access these Web services and the data. This will also allow maximum flexibility in the delivery of services from either national centres or the aggregated repository. Thus, on a case by case basis, it would be possible for the aggregated repository to either

directly provide the data and content or to act as a broker (aggregator) and request the data from one or more national centres for display through the portals.



7.2 Work package overview

**Work Package and Labour Effort Overview**

WP No <sup>10</sup>	Work package title	Lead AP No <sup>11</sup>	Start <sup>12</sup>	End <sup>13</sup>	Total PM <sup>14</sup>	Person months effort per work package per applicant													
						AP1	AP2	AP3	AP4	AP5	AP6	AP7	AP8	AP9	AP 10	AP 11	AP 12	AP 13	AP 14
1	Project coordination	1	M 0	M 36	<b>41</b>	30	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
2	Content management	1	M 0	M 36	<b>212</b>	66	13	6	2	0	1	7.5	12	4.5	12	7.5	3	4.5	9
3	Technological implementation	2	M 0	M 36	<b>203.4</b>	1	25	0	24	51	48	27	1	12	1	0	0	0	1
4	IPR issues	2	M 0	M 36	<b>18.6</b>	1	12	0.3	0	0	0.3	0.5	0.3	0.3	0.3	0.3	0.3	0.3	0.3
5	Dissemination & Exploitation	3	M 0	M 36	<b>49.4</b>	3	1	32	1	0.5	0.5	1	1	0.3	1	0.3	0.3	0.3	1
<b>Total</b>					<b>524.4</b>	<b>101</b>	<b>51.5</b>	<b>38.8</b>	<b>27.5</b>	<b>52</b>	<b>50.3</b>	<b>36.5</b>	<b>14.8</b>	<b>17.6</b>	<b>14.8</b>	<b>8.6</b>	<b>4.1</b>	<b>5.6</b>	<b>11.8</b>

WP No	Work package title	Person months effort per work package per applicant (continued)											
		AP 15	AP 16	AP 17	AP 18	AP 19	AP 20	AP 21	AP 22	AP 23	AP 24	AP 25	AP 26
1	Project coordination	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0	0	0
2	Content management	6	7.5	9	12	9	7.5	3	9	1	0	0	0
3	Technological implementation	0	6	0	0	4.5	0.3	0.3	0.3	1	0	0	0
4	IPR issues	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0	0	0	0
5	Dissemination & Exploitation	0.3	1	0.3	1	1	1	0.3	0.3	1	0	0	0
		<b>7.1</b>	<b>15.3</b>	<b>10.1</b>	<b>13.8</b>	<b>15.3</b>	<b>9.6</b>	<b>4.4</b>	<b>10.4</b>	<b>3.5</b>	<b>0</b>	<b>0</b>	<b>0</b>

<sup>10</sup> Work package number: WP 1 – WP n.

<sup>11</sup> Number of the applicant leading the work in this work package.

<sup>12</sup> Relative start date for the work in the specific work packages, month 0 marking the start of the project, and all other start dates being relative to this start date.

<sup>13</sup> Relative end date, month 0 marking the start of the project, and all ends dates being relative to this start date.

<sup>14</sup> The total number of person-months allocated to each work package.

### 7.3 Work package description

#### Work package Description

<b>Work package number :</b>	<b>1</b>	<b>Start date:</b>	<b>M1</b>	<b>End date:</b>	<b>M36</b>
<b>Work package title:</b>	<b>Project Coordination and management</b>				

#### Objectives

**Administrative objective:** Ensure adherence of the consortium to the rules, regulations, and financial guidelines of the *eContentplus* programme; establish the project in the European biodiversity community.

**Technical objective:** Implement the project as set out in the work plan; ensure exchange of information and communication between partners; ensure progress of the project; guarantee timely deliverables.

**QA objective:** Ensure verifiable progress of the project and be committed to high quality output that has tangible impact on *eContentplus* programme objectives.

#### Description of work

**Task 1.1 – Administrative coordination:** Resource planning (financial, personnel, material), monitoring and controlling; liaison between the Commission, consortium members and external experts; effective communication with the consortium members, Work Package leaders, the Commission, and interested external parties; coordination of meetings and progress reviews; production and consolidation of periodic external reports, including cost-statements; internal quality assurance; set-up of the Web-based project management portal; representing the project.

**Task 1.2 – Technical coordination:** Work-package and task coordination; work plan maintenance; monitoring of project progress and milestones; identification and trouble shooting of technical and organisational problems; timely production of deliverables; quality control against the technical and contractual aspects; coordination with EDL Foundation Office.

**Task 1.3 – Quality assurance:** Definition and communication of quality assurance procedures via project management portal, coordination of quality assurance process (internal as well as external review procedures for the various project results and deliverables).

**Task 1.4 – Networking and clustering activities:** Coordinate with relevant partner networks; contribute to the clustering activities of the EC.

#### (Inter-) Dependencies, milestones<sup>1</sup> and expected result

**Responsibility:** UBER (Project Coordinator) and all other consortium partners

**(Inter-) Dependencies:** WP1 is a horizontal work package accompanying the project from beginning to end.

**Milestone 1.1:** Project management portal including quality assurance procedures (M 3)

<sup>1</sup> Milestones are control points at which decisions are needed, for example concerning which of several technologies will be adopted as the basis for the next phase of the project.

**Deliverables**

- D 1.1 Progress Report 1 (M 6)
- D 1.2 Progress Report 2 and Annual Report 1 (M 12)
- D 1.3 Progress Report 3 (M 18)
- D 1.4 Progress Report 4 and Annual Report 2 (M 24)
- D 1.5 Progress Report 5 (M 30)
- D 1.6 Progress Report 6 and Final Report including Financial Statement (M 36)

**Work package Description**

<b>Work package number :</b>	<b>2</b>	<b>Start date:</b>	<b>M1</b>	<b>End date:</b>	<b>M36</b>
<b>Work package title:</b>	<b>Analysis of domain content and management of the content acquisition process</b>				

**Objectives**

**IT objective:** Establish bibliographic database systems, metadata repositories and Web-based content management systems.

**Management objective:** Ensure that all relevant biodiversity literature is listed to be scanned following a priority list; ensure that all content providers agree on the technical architecture of the project; ensure effective scanning in all content providing institutions; ensure linkage to EUROPEANA; ensure extension of the content providing network.

**Description of work**

**Task 2.1 – IT Development:** Establish a bibliographic database system and metadata repository for monographs and serials based on the Virtual Taxonomic Library developed in the EDIT project and experiences of BHL; develop a Web-database to support analysis of domain content and management of the scanning process (database of the taxonomic literature that indicates (a) the portion that is already available in digital form, (b) the portion that is in the process of being digitised, and (c) the portion for which plans have been created for digitisation).

**Task 2.2 – Analysis of domain content:** Establish a list of monographs and serials that are relevant for the biodiversity community; use of the Web-database to identify the distribution of this relevant literature in the libraries of the content providers; identify responsibilities for content contribution (which institution should provide the identified content under consideration of technical qualification, data standards and IPR issues).

**Task 2.3 – Management of the content acquisition process:** Identification of various content holder requirements; development of the Memorandum of Understanding; assist partners in implementation and evaluation of scanning operations; control duplicate scanning of literature (using results of task 2.2); coordinate with the EDLNet, BHL and national scanning projects to ensure that material scanned by BHL-Europe is available through these portals; discussion and distribution of data standards and specifications; addressing IPR issues in cooperation with WP4; take into account multicultural and multilingual aspects; work with individual donors and governments to facilitate the funding of the scanning; attracting new content providers.

**(Inter-) Dependencies, milestones<sup>1</sup> and expected result**

**Responsibility:** UBER (Work Package Leader) and all consortium content providers

**(Inter-) Dependencies:** WP2 will rely mainly on WP3 and WP4. A close cooperation with WP3 is important to ensure the content is accessible according to content holder requirements. A close cooperation with WP4 is important to legally integrate new content provider into the consortium. As content providers are the largest group in the consortium, WPL of WP2 does a lot of the communication work in support of WP1.

**Milestone 2.1:** Web-database for content management and collection analysis (M 6)

**Milestone 2.2:** Signed Memorandum of Understanding (M 6)

**Deliverables**

D 2.1 Bibliographic database system for serials and monographs (M 3)

D 2.2 Catalogue of content holder requirements (quality, quantity, accessibility, standards and specifications of content and metadata) (M 3)

D 2.3 Metadata and 20 million pages of biodiversity literature from 40 content providers delivered (M 36)

D 2.4 Delivery of the final revised best practice guidelines and standards (M 36)

---

<sup>1</sup> Milestones are control points at which decisions are needed, for example concerning which of several technologies will be adopted as the basis for the next phase of the project.

## Work package Description

<b>Work package number :</b>	3	<b>Start date:</b>	M1	<b>End date:</b>	M36
<b>Work package title:</b>	Technological implementation				

### Objectives

Management and coordination of technological development and associated standards to allow for a pan-Europe, distributed and multilingual BHL-Europe. The technological implementation will concentrate on the innovative application of proven technologies to deliver stable and sustainable solutions.

### Description of work

**Task 3.1 – Technological implementation (Overall Coordination):** Management of the technical development team; adaptation of EUROPEANA and BHL data model, workflow, harvesting procedure, standards, specifications; definition of own standards for images, metadata, harvesting according to partner requirements (consensus-building); liaise with scanning centres of the national initiatives for post-processing of content; development and adaptation of specific tools; implementation and adaptation of taxon finder and name recognition tools; improvement and implementation of OCR techniques.

**Task 3.2 – Technical integration with EUROPEANA, BHL and national platforms:** Ensure that the standards for data management and image formats are consistent with international requirements and specifically the EUROPEANA and BHL; to develop a distributed data model which will allow countries to retain control of their data, while enabling the material to be available through EUROPEANA, the BHL portal (with Taxonomic Intelligence tools), and any national portals; to build a prototype distributed data system for the German language material, and integrate with the BHL portal.

**Task 3.3 – Addressing distributed access and storage - long-term sustainability:** Develop a distributed access and storage system to enable national and international storage of the scanned materials; to develop the distributed storage system in such a way that long-term sustainability of the data is secured.

**Task 3.4 – Enabling BHL Portal access in European languages - interfaces, usability, and mobility:** Work with the EUROPEANA to create multiple language access to the BHL Portal; to create a prototype portal in German to allow access to the BHL Portal; to apply the model to the key European languages, enabling access to the BHL Portal throughout Europe.

### (Inter-) Dependencies, milestones<sup>1</sup> and expected result

**Responsibility:** WP3 will be led by NHM London. It is important that this work package is led by a partner institution that has capacity and experience in delivering this type of project. The NHM's leadership role in BHL and EOL as well as its contribution to other related projects makes it a suitable candidate.

**(Inter-) Dependencies:** WP3 will rely on the outputs and all the other work packages, but particularly WP1, WP2 and the online elements of WP5.

Each sub-work package in WP3 will have a specific structure, starting with production of an options appraisal and assessment papers that will review standards and technologies. This will enable the technology development team/board to agree standards policy or implementation plans

<sup>1</sup> Milestones are control points at which decisions are needed, for example concerning which of several technologies will be adopted as the basis for the next phase of the project.

for specific products under each heading.

**Milestone 3.1:** All sub-work packages produce options appraisals for all products/work streams and detailed implementation plans agreed for years 2 and 3. This would include the issuing of all standards, data models, technology standards, preferred technologies etc. required for implementation. (M 6)

**Milestone 3.2:** Technology review (M 24)

#### **Deliverables**

D 3.1 Deliver composition of Technology Management Board (M 3)

D 3.2 Plan for rolling out agreed standards, best practice and system components (M 9)

D 3.3 Implement plans for all products in WP3, incl. data models, technology standards etc. (M 12)

D 3.4 Release German prototype (M 18)

D 3.5 Key components defined in output of D 3.4 (e.g. BHL-Europe Portal, OCR demonstrators, distributed storage model) (M 24)

D 3.6 Sustainability policy for continuation of service. (e.g. hosting, future development, helpdesk provision for service users/content providers etc.) (M 30)

D 3.7 Live BHL–Europe with distributed storage and management and appropriate tools for the continued development of services and ingress of multilingual content (M 36).



## Work package Description

<b>Work package number :</b>	4	<b>Start date:</b>	M1	<b>End date:</b>	M36
<b>Work package title:</b>	<b>Intellectual Property Rights</b>				

### Objectives

Management and coordination of the intellectual property rights (IPR) framework for BHL-Europe and agreements with Rights Holders.

### Description of work

**Task 4.1 – IPR Framework:** Establish IPR working documents - including best practice guide, due diligence guide, pro forma agreements, and process for formally agreeing rights management with rights holders; align the approach with EUROPEANA and BHL.

**Task 4.2 – IPR Agreements with Data Providers:** Complete formal IPR agreements with data providers identified in WP2; keep records of all agreements secure.

**Task 4.3 – IPR Agreements on projects results and outcomes:** Develop IPR framework for long-term sustainability and long-term access to the digitised content of BHL-Europe.

### (Inter-) Dependencies, milestones<sup>1</sup> and expected result

**Responsibility:** WP4 will be led by Participant 2 – the NHM, London. Participant 2 has a dedicated IPR Officer and access to the advice of a law firm (Farrer and Co.) who are experts on IPR.

**(Inter-) Dependencies:** WP2 will rely on WP4 to ensure that material made available through WP3 has appropriate IPR agreements in place. WP2 will identify potential data providers with whom we intend to establish IPR agreements. Development of a sustainability framework is dependent on contributions from WP2, WP3, and WP5.

**Milestone 4.1:** A working agreement on IPR will be in place with EUROPEANA. (M 9)

### Deliverables

D 4.1 Delivery of IPR working documents, including best practice guide, due diligence guide, pro forma agreements and process for formally agreeing rights management with rights holders (M 9)

D 4.2 Complete signed agreements with rights holders to enable material to be used in the BHL or EUROPEANA (M 36)

D 4.3 Deliver IPR framework to support long-term access and sustainability of the digitised material (M 36)

<sup>1</sup> Milestones are control points at which decisions are needed, for example concerning which of several technologies will be adopted as the basis for the next phase of the project.

## Work package Description

<b>Work package number :</b>	<b>5</b>	<b>Start date:</b>	<b>M1</b>	<b>End date:</b>	<b>M36</b>
<b>Work package title:</b>	<b>Dissemination, Exploitation and Evaluation</b>				

### Objectives

**Dissemination objective:** Develop dissemination strategy; raise awareness, understanding and action of the project among the community and stakeholders in EU member states; ensure effective dissemination of project goals and results to the target users; ensure good communication within the European scientific community

**Presentation and demonstration objective:** Ensure dissemination of project results at conferences, public events and among the networks of the consortium members

**Exploitation objective:** Develop sustainability strategy for the results of BHL-Europe; ensure implementation of project results in other projects, initiatives, institutions, and countries

**Evaluation objective:** Monitor the level of use of BHL-Europe; survey the users of the Web portal

### Description of work

**Task 5.1 – Dissemination plan & activities:** Develop promotional concept; develop identity and corporate design; plan, design and publish BHL-Europe Web site for internal and external communication with publication of all relevant results and links to partner networks (with regular updates); design and produce BHL-Europe promotion kit (flyers, posters, presentations - with regular updates); develop newsletter and mailing list to facilitate communication of new developments (internal and external); identification and analysis of stakeholders, definition of their benefits from BHL-Europe.

**Task 5.2 – Demonstration plan & activities:** Database of conferences and public events relevant for BHL-Europe; identify consortium members to present the BHL-Europe promotion kit at selected events; provide the promotion kit to consortium members to disseminate BHL-Europe within their networks.

**Task 5.3 – Exploitation plan & activities:** Establish BHL-Europe Communications Working Group; investigate services that help the sustainability of project results; develop a business plan for long term sustainability with WP2 and WP3; identify related networks or organisations that might be interested in implementation of BHL-Europe results.

**Task 5.4 – Evaluation plan & activities:** Put in place Web-based evaluation tools to survey users (to determine target group of the users, country of the users, most interesting content, etc.); develop online questionnaires to identify user requirements, preferences, experiences, benefits, and un-met needs; document case studies of non-science usage of BHL-Europe content.

### (Inter-) Dependencies, milestones<sup>1</sup> and expected result

**Responsibility:** WP5 will be led by National Museum Prague. This museum has an experienced team for communications, public relations, and related issues.

**(Inter-) Dependencies:** WP5 will rely mainly on WP3 to start the exploitation and evaluation activities.

<sup>1</sup> Milestones are control points at which decisions are needed, for example concerning which of several technologies will be adopted as the basis for the next phase of the project.

**Deliverables**

- D 5.1 First version of Web site and promotion kit (M 3)
- D 5.2 BHL-Europe newsletter and mailing list (M 3)
- D 5.3 Database of relevant conferences/events and ownership for BHL-Europe presentations (M 3)
- D 5.4 Deliver composition of BHL-Europe Communications Working Group (M 3)
- D 5.5 BHL-Europe dissemination plan (M 6)
- D 5.6 BHL-Europe Web site and promotion kit, final version of (M 9)
- D 5.7 Online questionnaires for user survey (M 12)
- D 5.8 First user evaluation report (M 15)
- D 5.9 Second user evaluation report (M 30)
- D 5.10 Business plan for long-term sustainability (M 36)
- D 5.11 BHL-Europe presentation, final version (M 36)
- D 5.12 Implementation of results of BHL-Europe in other projects (M 36)

## 7.4 Deliverables List

<b>Deliverables List</b>				
<b>Deliverable No<sup>1</sup></b>	<b>Deliverable title</b>	<b>Delivery date<sup>2</sup></b>	<b>Nature<sup>3</sup></b>	<b>Dissemination level<sup>4</sup></b>
1	D 2.1 Bibliographic database system for serials and monographs	M 3	P	PP
2	D 2.2 Catalogue of content holder requirements (quality, quantity, accessibility, standards and specifications of content and metadata)	M 3	R	CO
3	D 3.1 Deliver composition of Technology Management Board	M 3	O	CO
4	D 5.1 First version of Web site and promotion kit	M 3	D	PU
5	D 5.2 BHL-Europe newsletter and mailing list	M 3	D	PU
6	D 5.3 Database of relevant conferences / events and responsibilities for BHL-Europe presentations	M 3	O	CO
7	D 5.4 Deliver composition of BHL-Europe Communications Working Group	M 3	O	CO
8	D 1.1 Progress Report 1	M 6	R	CO
9	D 5.5 BHL-Europe dissemination plan	M 6	R	CO
10	D 3.2 Plan for rolling out agreed standards, best practice and system components	M 9	R	CO
11	D 4.1 Delivery of IPR working documents, including best practice guide, due diligence guide, pro forma agreements and process for formally agreeing rights management with rights holders	M 9	R	CO
12	D 5.6 BHL-Europe Web site and promotion kit, final version of	M 9	D	PU
13	D 1.2 Progress Report 2 and Annual Report 1	M 12	R	PU
14	D 3.3 Implement plans for all products in WP3, including data models, technology standards etc.	M 12	O	CO
15	D 5.7 Online questionnaires for user survey	M 12	O	PU
16	D 5.8 First user evaluation report	M 15	R	PP
17	D 1.3 Progress Report 3	M 18	R	CO

<sup>1</sup> Deliverable numbers in order of delivery dates: D1 – Dn. Deliverable numbers must indicate which work package they relate to, e.g. D2.1 for the first deliverable from work package 2).

<sup>2</sup> Month in which the deliverables will be available. Month 0 marking the start of the project, and all delivery dates being relative to this start date.

<sup>3</sup> Please indicate the nature of the deliverable using one of the following codes:

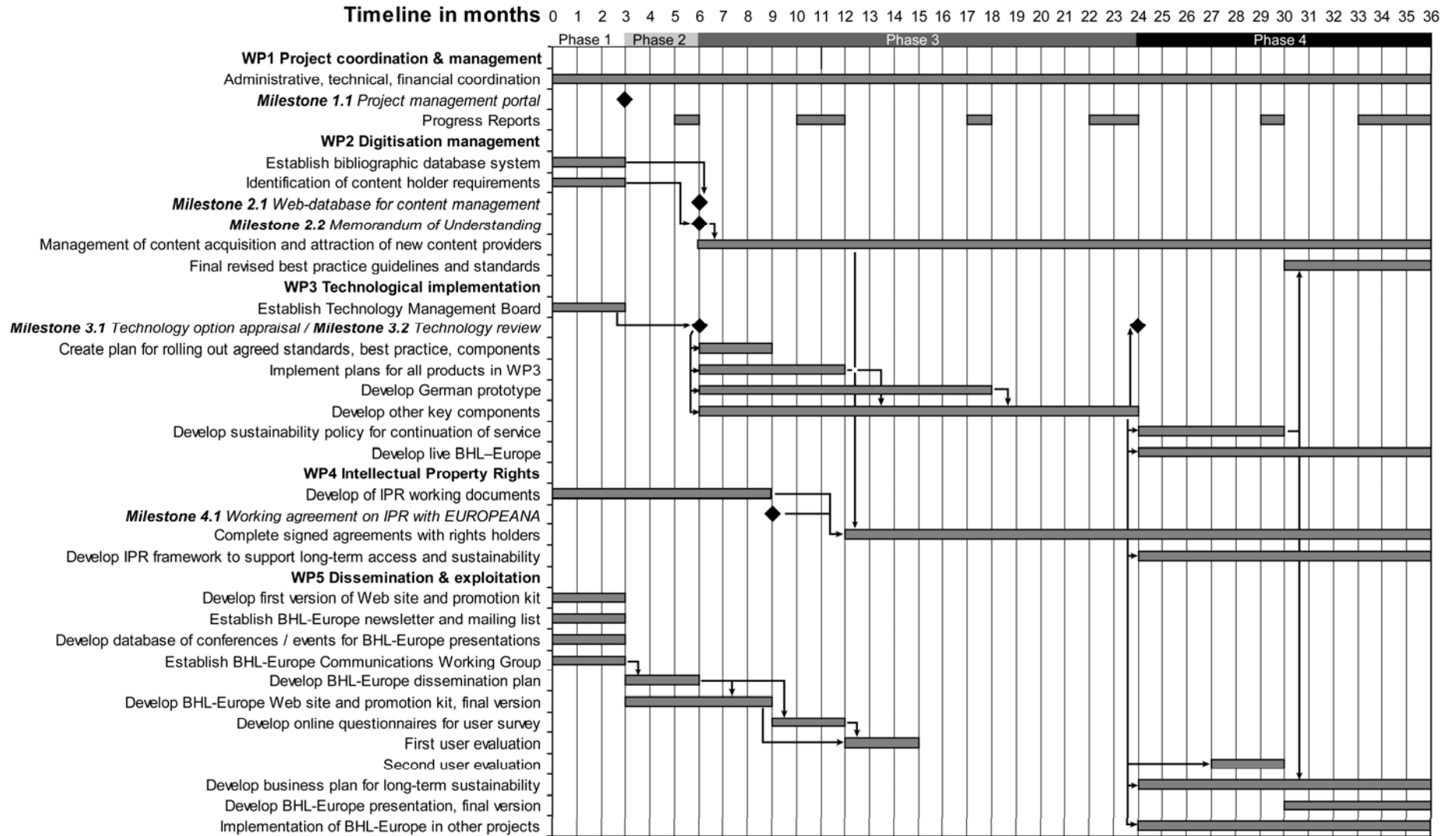
**R** = Report  
**P** = Service/Product  
**D** = Demonstrator/Prototype  
**O** = Other

<sup>4</sup> Please indicate the dissemination level using one of the following codes:

**PU** = Public  
**PP** = Restricted to other programme participants (including Commission services and project reviewers).  
**CO** = Confidential, only for members of the consortium (including Commission services and project reviewers).

18	D 3.4 Release German prototype	M 18	D	PU
19	D 1.4 Progress Report 4 and Annual Report 2	M 24	R	PU
20	D 3.5 Key components defined in output of D 3.4 (e.g. BHL-Europe Portal, OCR demonstrators, distributed storage model)	M 24	P	PU
21	D 1.5 Progress Report 5	M 30	R	CO
22	D 3.6 Sustainability policy for continuation of service. (e.g. hosting, future development, helpdesk provision for service users/content providers etc.)	M 30	R	CO
23	D 5.9 Second user evaluation report	M 30	R	PP
24	D 1.6 Progress Report 6 and Final Report including Financial Statement	M 36	R	CO
25	D 2.3 Metadata and 20 million pages of biodiversity literature from 40 content providers delivered (M 36)	M 36	P	PU
26	D 2.4 Delivery of the final revised best practice guidelines and standards	M 36	R	PP
27	D 3.7 Live BHL–Europe with distributed storage and management and appropriate tools for the continued development of services and ingress of multilingual content	M 36	P	PU
28	D 4.2 Complete signed agreements with rights holders to enable material to be used in the BHL or EUROPEANA	M 36	P	CO
29	D 4.3 Deliver IPR framework to support long-term access and sustainability of the digitised material	M 36	R	CO
30	D 5.10 Business plan for long-term sustainability	M 36	R	CO
31	D 5.11 BHL-Europe presentation, final version	M 36	D	PU
32	D 5.12 Implementation of results of BHL-Europe in other projects	M 36	P	PP

7.5 Project plan

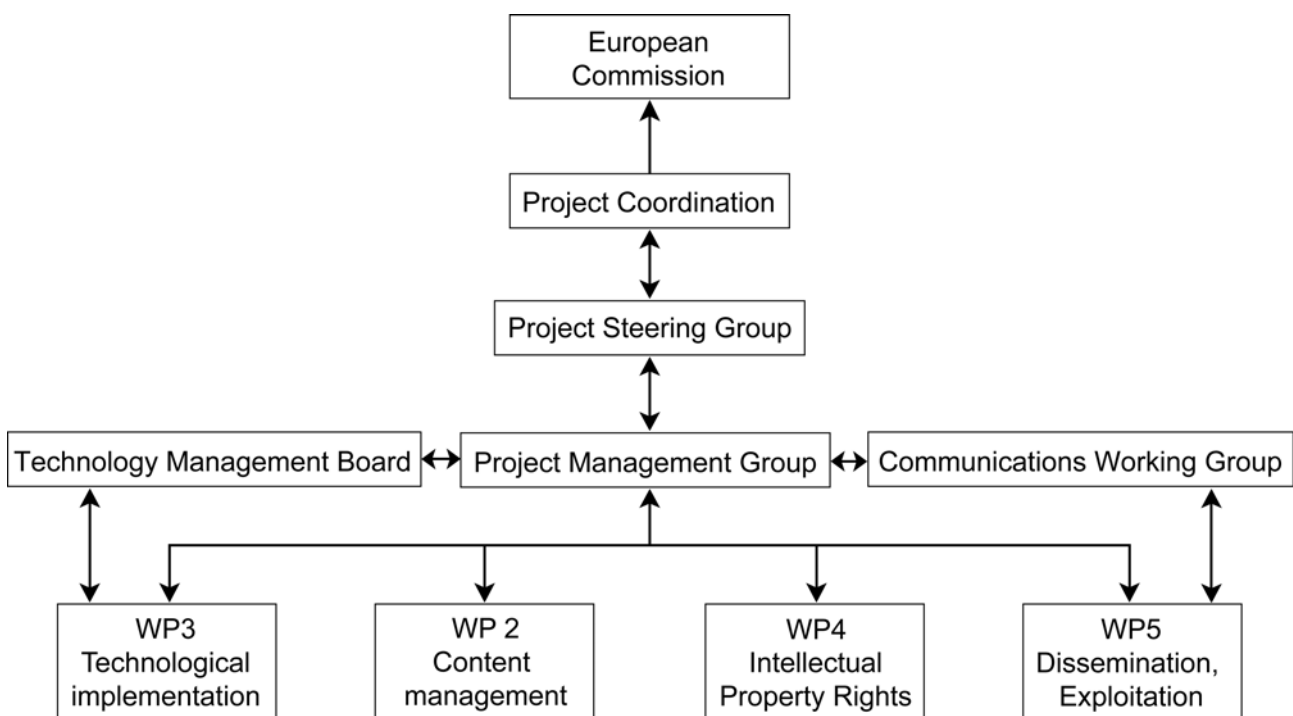


## 8 Project management

### 8.1 Project Management Structure

BHL-Europe includes a considerable number of partners and the consortium will grow more in the lifetime of the project. To ensure an effective management of the project, working documents and decisions will be prepared at different levels. The BHL-Europe management structure consists of five different management levels:

- The project coordination (PCO)
- The Project Steering Group (PSG)
- The Project Management Group (PMG)
- The Technology Management Board (TMB) and Communications Working Group (CWG)
- The Work Packages (WP)



### 8.2 Project Coordination (PCO)

BHL-Europe will be coordinated by Dr. Henning Scholz based at the Museum für Naturkunde of the Humboldt-University of Berlin (UBER). The museum has been involved in the management of the EU I3 SYNTHESYS project since 2004. SYNTHESYS aims to create an integrated European infrastructure for researchers in the natural sciences. In addition, the museum is involved in the development of numerous national and international projects funded by the EU (EDIT), BMBF (BIOTA, GBIF), BMU (Acoustic monitoring of breeding birds at wetlands in the region of the River Peene), DFG (Graduate Research Program 503, Research Group 533 and 736). As part of the Humboldt-University, the coordinator also benefits from the vast experience of the Humboldt-University in managing EU projects, including DoD (Digitisation on Demand) and ALEAR (Artificial Language Evolution on Autonomous Robots). In its role as coordinator, UBER will be responsible to ensure that the project is run in accordance with the work plan and the overall contract. UBER will be responsible for all administrative and technical matters of BHL-Europe.

The administrative coordination includes:

- resource planning (financial, personnel, material), monitoring and controlling
- liaison between the Commission, consortium members and external experts
- effective communication with the consortium members, Work Package leaders, EU officials and interested external parties
- coordination of meetings and progress reviews
- production and consolidation of periodic external reports, including cost-statements
- internal quality assurance

The technical coordination includes:

- work package and task coordination
- work plan maintenance
- monitoring of project progress and milestones
- identification and trouble shooting of technical and organisational problems
- timely production of deliverables
- quality control of technical and contractual aspects

### **8.3 Project Steering Group (PSG)**

The Project Steering Group will consist of one representative from each partner of the BHL-Europe consortium, including the project coordinator:

- Henning Scholz (UBER, Project Coordinator) – Chair of PSG
- Graham Higley (NHM), Jiri Kvacek (NMP), Jill Cousins (EDL), Walter Koch (AIT), Roger Essoh (ATOS), Walter Berendsohn (FUB-BGBM), Francisco Welter-Schultes (UGOE), Ernst Vitek (NHMW), Fritz Gusenleitner (OOELM), Laszlo Peregovits (HNHM), Robert Turlej (MIZPAS), Henning Knudsen (SNM), Kees Hendriks (NAT), Elmar Robbrecht (NBGB), Patricia Mergen (RMCA), Patrick Grootaert (RBINS), Hervé Colinmaire (BnF), Michelle Lenoir (MNHN), Antonio Valdecasas (CSIC), Luca Bartolozzi (MSN), Jane Hutcheon (RBGE), Frank Bisby (Sp2000), Philippa Scoones (Wiley), Thomas Garnett (SIL), Christopher Freeland (MOBOT).

The PSG will be responsible for providing direction and the policy making role for the project. It will also be the ultimate body to resolve with conflicts and disputes that may arise in the consortium. The PSG will also be responsible for discussing, communicating, and deciding on any contractual amendments and issues concerning the European Commission, including changes in funding allocation or work plan.

The PSG will also decide on the overall communication and dissemination strategy of the project. The group will convene twice a year in regular meetings or through e-meetings, and will call additional meetings if needed.

Decisions of the PSG will be often by consensus but when necessary decisions will taken by a straight majority vote with the chair having a casting vote in the event of a tie.

### **8.4 Project Management Group (PMG)**

The Project Management Groups is formed by the Work Package leaders (WPL) of the BHL-Europe project. A Project Manager from the European Digital Library Foundation and the Chair of the BHL (Graham Higley) will attend the PMG to assist the operational collaboration with EUROPEANA and BHL. Its prime tasks are the overall operational management of BHL-Europe, which includes supervising and monitoring the actual work carried out in the work packages, reviewing deliverables, and ensuring that the (technical) objectives and goals are fulfilled. The



PMG serves as the first instance for resolving conflicts in the consortium (the PSG only gets involved, if disputes cannot be settled on the level of the PMG). Depending on the issues to be discussed, the PMG may be temporarily enlarged by additional key personnel working on particular work package tasks. The group will convene four times in the first year and at least twice in each of the following years in ordinary meetings or through e-meetings, and will call additional meetings if needed.

Decisions of the PMG will be often by consensus but when necessary decisions will be taken by a straight majority vote of the full members with the chair having a casting vote in the event of a tie.

Members of the PMG are:

- Henning Scholz (UBER, Project Coordinator) – Chair of PMG
- Michael Ohl (UBER, WP2)
- Paul Richards (NHM, WP3)
- Nancy Chillingworth (NHM, WP4)
- Jiri Kvacek (NMP, WP5)
- Graham Higley (NHM, Chair of BHL)
- Jill Cousins (EDL)

### **8.5 Technology Management Board (TMB)**

The Technology Management Board will be responsible for agreeing on standards policy and implementation plans. The members of the TMB will also revise the results of the project and help in the technical evaluation of project's outcomes. In addition to the members of the WP3 doing the actual work (NHM, AIT, ATOS, EDL, FUB-BGBM, NHMW), several other partners have a strong IT and digital library expertise that is helpful in the preparation of work. These partners include UBER (Susanne Dobratz), OOELM (Michael Malicky), NAT (Berry van der Hoorn), RMCA (Danny Meirte), and MOBOT (Christopher Freeland) and are foreseen as potential contributors to TMB. In addition to the personal meetings they will have e-meetings several times a year. They will regularly report to the PMG through the chair of the TMB, which is the leader of WP3 (Paul Richards, NHM).

### **8.6 Communications Working Group (CWG)**

The BHL-Europe Communications Working Group will be responsible for revising and discussing dissemination strategies and agreeing on related activities. In addition to the leading institution for all activities related to dissemination and exploitation (NMP), members of other consortium partners are invited to attend the board. Although consensus building is the main aim of the activities, discussions with all partners involved might be difficult because of the large consortium. To ensure an effective management at this level, working documents will be developed first in the CWG before they are revised in the PMG and PSG. The final selection of CWG members is done in the first months of the project. At the moment, members from UBER, NHM, EDL, RMCA, NAT, MNHN, and SIL are foreseen as potential contributors to the CWG. In addition to the personal meetings they will have e-meetings several times a year. They will regularly report to the PMG through the chair of the CWG, which is the leader of WP5 (Jiri Kvacek, NMP).

### **8.7 Work Packages (WP)**

The Work Package leaders are responsible for coordinating and managing their work packages. In particular, this includes the detailed planning of the work that needs to be carried out, the coordination and division of work among the project partners, the responsibility to ensure that enough resources are allocated to get the work done, and finally, the delivery of the result within the agreed upon schedule. The Work Package leaders will report on a regular basis to the PMG and the

PCO on work progress, and give due notice about any potential delay to achieving the deliverables set out in the work plan. In addition, they are also responsible for scheduling meetings of the Work Package members and relevant members of other work packages to discuss the technical issues of their work packages.

### **8.8 Means for communication, monitoring and assuring quality**

The primary method for information exchange between the consortium members will be through project meetings at various management levels (Project Steering Group, Project Management Group, and Work Packages). Meetings will be announced and scheduled at least four weeks in advance; virtual meetings can be scheduled on shorter notice. Agendas, proposed resolutions, and working documents will be made available to all attendees at least one week before the meeting. For meetings of the PSG and PMG, responsibility for preparing the meeting rests with the project coordinator. For Work Package meetings, the Work Package leader who calls the meeting will be responsible for preparation. Draft minutes of a meeting will be issued within two working days after the meeting.

An additional platform for communication between the consortium members will be a Web-based project management portal with restricted access for consortium members which will be established within three months after the start of the project. On this platform, the PCO provide management information and guidelines, the work plan, templates for reports and the quality assurance procedure. The PCO is responsible for regular updates (every three months) of the management information and to provide a status report as an internal review of the project. Consortium members may discuss specific problems, get all relevant information on upcoming meetings and new developments, and publish drafts of deliverables for internal review. Once deliverables with public status have been approved, they will be published on the BHL-Europe Web site in an open public area. The project management portal also contains the documentation of internal communication throughout the life-time of the project. The portal will also have specific tools and functionalities that facilitate and optimises book-keeping. In addition to the portal, a newsletter will be established to inform the members of the consortium, and other interested parties, about important developments in BHL-Europe. Various mailing lists adapted to the project management structure are provided to every member of the consortium in order to facilitate communication within larger groups of members.

As part of the quality assurance procedure, each partner has to deliver an internal report every three months to the PCO. The PCO will summarise these reports to inform the PSG on the project status on a quarterly basis. For all internal reporting, the Project Coordinator will provide the necessary templates through the BHL-Europe Web site.

Reporting to the European Commission will be done according to contractual requirements, providing a status report on activities as well as finances. All partners report to the PCO who will send an aggregated report to the EC Project Officer. The templates for official reporting are provided by the European Commission.

### **8.9 Means for resolving conflicts and risk management**

The Work Package leaders will immediately inform the PCO if any problems in achieving their technical objective or if any potential conflict situations arise. As already mentioned above, the first opportunity to resolve disputes is in the PMG. If conflicts cannot be resolved there, they will be brought into the PSG who either can settle the conflict or suggest/decide strategies to further deal with the conflict.

Risk management requires risk identification, control, and recording of risks, highlighting of the consequences and the appropriate management actions. Risk management is a balance of judgement so that the risks are minimised without over-emphasising the potential problems. Controlling the

risks will help us to manage our project to achieve properly the objectives on time and to budget. Risk management will be an integral part of the BHL-Europe lifecycle management process. In this sense, risk assessment methods will be applied, where needed, in order to minimise possible deviations from the expected results and schedule.

With regards to risk management, the PMG will have the responsibility to review and evaluate different situations which may lead to some kind of risk. To facilitate this procedure, the regular reporting via the project management portal will be used to consider and evaluate risk factors. Work package leaders will report apparent and imminent risks immediately, and the PMG will decide upon and immediately carry out remedial actions.

One important operational risk identified is related to the consensus building in order to sign the BHL-Europe Memorandum of Understanding. Some content holders might have particular requirements that are not compatible with the requirements of most other partners. However, the technologies we are intending to use are proven to be flexible enough (e.g. FEDORA) to deal with various partner requirements to accomplish the aim of BHL-Europe. Otherwise, we are anticipating no major science and technology risks. We focus on the integration of existing and proven technologies to deliver stable and sustainable solutions.

## **9 Description of consortium and key personnel**

### **9.1 Description of partners and key personnel**

#### **9.1.1 Humboldt-Universität zu Berlin (Museum für Naturkunde) (UBER)**

The Museum für Naturkunde (MfN) is part of the Humboldt-University in Berlin, one of the most prestigious and eminent universities in Germany. The MfN is deeply embedded in the intellectual environment of the university since its foundation in 1810. With more than 30,000,000 objects, the Museum für Naturkunde is the largest natural history museum in Germany and among the largest museums worldwide. The library of the MfN holds 175,000 items and about 1,000 currently subscribed journal titles in the fields of zoology, palaeontology, and mineralogy, with the oldest book dating back to 1473. The museum has been involved in the management of the EU I3 SYNTHESYS project since 2004. In addition, the museum is involved in the development of numerous national and international projects funded by the EU (e.g. EDIT), BMBF (Biota, GBIF), BMU, and DFG (e.g. Graduate Research Program 503). There is a dedicated grant administration team at UBER who assist grant holders with the delivery of the project including contractual and financial management and reporting. The University Library and the Computer and Media Services of UBER have cooperate strongly and have carried out several projects within the fields of e-publishing, digital preservation and the digital library. Both institutions have been involved in several EU projects during the last 5 years: Open Archives Forum (EU-IST Programme); reUSE (eContentplus project); Digitisation-on-Demand (eTEN project).

#### **Dr. Henning Scholz**

Henning Scholz holds a PhD in Palaeontology. His broad biodiversity expertise comes from his scientific work on the zoology and palaeontology of marine and freshwater molluscs and amphibians. Since 2002 he has worked for MfN in various functions. From 2002 to 2004 he was project manager of the Graduate Research Program 503. Since 2004 he mainly works as a curator of Invertebrate Palaeontology and assistant of the Head of the Department of Collections. In 2007 he was part of the management team responsible for the development of the new permanent exhibition of the MfN.

**Manja Voß**

Manja Voß holds a MSc in Geology-Palaeontology and is DE-TAF administrative officer (SYNTHESYS) at MfN since 2005. She is responsible for day-to-day administration, and has extensive experience with the coordination of the project-oriented workflows (e.g. planning of conferences, communication and arrangement with international scientific visitors, qualitative and quantitative assessment of collection data for evaluation purposes).

**Dr. Michael Ohl – Head of the Entomological Collections and the Library Committee**

Besides being a biologist, Michael Ohl has also studied philosophy and history of sciences at the universities of Kiel and Göttingen, which makes his work interests closely connected to libraries and historical books. He is also engaged in building up the German initiative to scan German biodiversity literature.

**Prof. Dr. Peter Schirmbacher – Director of the Computer and Media Centre at UBER**

Peter Schirmbacher graduated in economic science and holds a PhD in applied information science. Since 2006 he is additionally Professor for Information Management at the Institute of Library and Information Science of UBER. He led several projects in the area of "electronic publishing", multimedia applications, "e-Learning" and "e-Science". His current or recently concluded projects are OA-Netzwerk, HyperImage, e-kokon; reUSE; SCOPE.

**Susanne Dobratz**

Susanne Dobratz studied Informatics at the Technical University of Berlin, with emphasis on artificial intelligence, machine learning and database systems. She was project coordinator for the Electronic Publication of Dissertations project (DiDi) of UBER. She is currently working on the Proprint – Printing on Demand Project. She set up the Open Archives compatible server for UBER.

**9.1.2 Natural History Museum (NHM)**

The Natural History Museum is one of the world's great museums, with over 3,300,000 visitors and 11,000,000 online visitors per year. The NHM is also an international leader in the scientific study of the natural world. NHM has a strong track-record in EC funded research and training. It currently leads the €13m EU I3 Programme SYNTHESYS project, which provides access and training in 20 natural history museums and herbaria (26,000 days in total to collections in 11 countries). NHM is currently in negotiation on nine FPVII projects across a range of research disciplines. In addition, NHM has led training networks and infrastructure projects in the last three frameworks; currently leads a FP6 RTN project, ORIGINS and is a partner in a further 15 live FP6-funded. The NHM Library has the largest collection of natural history materials in the world, with over 1,000,000 books (from 1469 onwards), 25,000 journal titles, and 600,000 works of art. The NHM is a leading participant in the BHL and an active player in the EOL Project. The NHM is committed to making its store of information freely available to as many people as possible. NHM has a large number of staff members with good background in Library Management Systems and strong IT skills including programming and database creation sharing their expertise with BHL-Europe.

**Graham Higley – Head of Library and Information Services**

Graham Higley's experience of EU and global project management is extensive, and includes multi-million Euro IT programmes and large building projects. He is the Coordinator for the I3 FPVI funded SYNTHESYS contract, the Chair of the BHL and on the Steering Committee for the EOL Project. He is also a member of Science Group (which sets policy for all NHM Science) and Chair of the NHM Information Strategy Group. He is ideally-placed to provide the linkage between the BHL-Europe project and the BHL.

**Paul Richards – Head of ICT**

Before becoming the Head of ICT at NHM in 2005, Paul Richards worked in the UK university sector and the UK National Health Service. He has post graduate qualifications in IT and relevant

professional memberships and qualifications including in project management. He has successfully delivered IT infrastructure projects valued up to £3 million. His previous technical roles focused on database technologies and system development.

**Chris Sleep** – Head of Infrastructure and Systems

Chris Sleep has worked at the Museum for over 10 years in various technical roles and has been head of systems and infrastructure for the last 3 years. He has extensive experience in the UK academic environment and has support a number of similar projects including Synthesis and EDIT. His expertise includes system integration and system design.

**Nancy Chillingworth**

She is an experienced IPR professional and will be joining the NHM on 1 July 2008.

**9.1.3 Narodni muzeum (NMP)**

The National Museum is a public scientific institution which systematically enriches its collections consisting of objects of natural and historical sciences all over the world, with particular interest to the Czech Republic. It conducts research in various fields of natural and historical sciences and has a large exhibit activity. The National Museum is the most distinguished and the largest museum in the Czech Republic. It consists of five professional institutions: Natural History Museum, Historical Museum, The Náprstek Museum of Asian, African and American Cultures, Czech Museum of Music, National Museum Library. At present the National Museum houses almost 20 million items from the area of natural history, history, archaeology, arts, music and librarianship.

**Dr. Jiří Kvaček** – Head of the Department of Palaeontology

Jiří Kvaček studied Geology in Prague and holds a PhD in Palaeontology. His scientific work is focused on palaeobotany. From 2005 to 2008 he was science director of NMP. He was involved in management of several scientific projects in palaeontology, a project of Scientists night and three exhibitions. In 2007 he organised the management team responsible for the development of the new permanent exhibition of NMP.

**Dr. Pavel Dousa** – Head of the Centre for Presentation of Cultural Heritage

Pavel Dousa holds a PhD in historical sciences and MA in history and museology. He is board member of the Museology committee, the Museum education committee of Czech Association of Museum and Galleries, and the National Jury of Gold Coat of arms competition for the best City Web.

**PhD Richard Šípek** – curator of the old prints collection in the library of NMP

Richard Šípek studied bibliology at The Institute of Librarianship and Information Sciences (Charles University - Prague). He runs and participates in several projects connected with research of the historical private book collections (Otto Jr. of Nostitz, Ferdinand II of Tyrol). In the meantime, he is a member of The Institute of Librarianship and Information Sciences.

**Pavel Stastný**

Pavel Stastný passed High school of design in Prague. He worked in various companies and state organisations. He is specialised in production of PR materials for state (e.g. Academy of Sciences of the Czech Republic) and private subjects (Datart). He developed corporate design for journals of the NMP, produced numerous exhibits (e.g. Architecture Week, Prague 2007). He was involved in several design projects for web portal Seznam.cz. He also works as an independent designer and exhibited his design in London, Paris, New York and Mexico.

**Petr Zrustek and Lukas Zvolanek**

Both are IT specialists at NMP and are involved in the development of IT services and Web programming and development.

#### 9.1.4 EDL Foundation (EDL)

The Stichting European Digital Library (EDL Foundation) is a cross domain foundation, under Dutch law, set up for the purpose of fostering collaboration between Museums, Archives, Libraries, and Audiovisual Collections in Europe. It aims to provide access to Europe's cultural heritage by facilitating formal agreement across museums, archives, audio-visual archives, and libraries on how to cooperate in the delivery and sustainability of a joint portal. It also provides a legal framework for use by EU funded projects to bring their research or content into the EUROPEANA. Its current Board of Participants is made up of pan-European Associations from the 4 sectors able to represent and mobilise their members to contribute and form part of EU funded projects aimed at realising a European Digital Library. These are EURBICA, FIAT, ACE, EMF, ICOM Europe, CENL, CERL, LIBER, MICHAEL, the Koninklijke Bibliotheek, INA, the Bundesarchiv, and the BnF.

##### **Dr. Jill Cousins**

Jill Cousins is the Executive Director of EDL Foundation and the Programme Director for The European Library (TEL) and EUROPEANA. She was responsible for creating the operational service The European Library. Its success has led to the EU giving their strategic backing to TEL for the creation of the EDL. She has a strong Web publishing background, having worked for VNU as their European Business Development Director and then in scholarly publishing, working for Blackwell Publishing.

##### **Catherine Lupovici**

Catherine Lupovici has many years of experience in digital libraries, being Director of the Digital Library Department at the BnF prior to joining EUROPEANA Net Office. Catherine was involved in the development of the Web archiving services in the BnF, in the development of the trusted digital repository and in the contribution of BnF to the definition of what a European Digital Library might be through the French EUROPEANA maquette and prototype as well as in the Gallica project. She also worked in Jouve SA French printed company involved in data capture and electronic publishing where she led several European Research and Development projects bringing together libraries, publishers, and research laboratories specialised in information technologies.

##### **Olaf D. Janssen**

Olaf D. Janssen is a project & account manager for The European Library (TEL). In 2004 he was asked to coordinate the technical & editorial development of the project. Before, he was editor in chief for Gabriel, the precursor of TEL. He has been involved in a number of EU-funded projects to expand and improve The European Library. Currently he is a project manager for two European projects, TELplus (2007-2009) and FUMAGABA (2008-2009).

#### 9.1.5 Angewandte Informationstechnik Forschungsgesellschaft mbH (AIT)

AIT is an Austrian software and research company founded in 1979. Research work is done primarily in the field of information management (e.g. distributed databases, collection management and knowledge engineering). It is carried out within the EC (e.g. IST, Ten-Telecom etc.) action lines or on national and regional level. Among some of the research projects that AIT cooperated with or co-ordinated count MOSAIC (Museums Over States and virtual Culture; TEN-Telecom), COVAX (Contemporary Culture Virtual Archives in XML; IST-Programme), REGNET (Cultural Heritage in REGIONAL NETWORKS), Media.Alp (Setting up an integrated communication platform for achieving a cultural community in the Alpine space; Interreg Alpine Space Programme), and DISMARC (DIScovering Music ARCHives; eContentplus). Within DISMARC OAI technology and protocols are used to create a common catalogue of distributed archives metadata. At regional level AIT is the technical provider for the DIS project, where a virtual content catalogue for museums, archives, libraries and other content institutions is being created based on OAI technology and international standards (Dublin Core). Further research work focuses on the provision of Web services for the cultural heritage domain. These Web services are based on the SKOS vocabulary

and first implementations concern the provision of vocabulary Web services as support process to the primary procedure "Cataloguing" (implemented standards: ANSI/NISO Z39.19, ISO 2788:1986, ISO-5964).

**Prof. Dr. Walter Koch** – Director of AIT Ltd.

Walter Koch is chairperson of the CSC Europe EEIG and head of the Steinbeis Transferzentrum (IMCHI – Information Management and Cultural Heritage Informatics). He is consultant and contractor to various national and international organisations (e.g. UNESCO, DFG, EC, ESA, Austrian ministries, Graz municipality) and member of several national and international scientific associations (e.g. ICOM, ONORM, VÖB). His 10 years project experience includes: bibliographic information, information systems, IT-management; EU Projects (TAP, Raphael, TenTelecom, IST, Interreg IIIB, eContentplus): COVAX, OpenHeritage, CULTIVATE, CIMI, MEDIA.ALP.

**Odo Benda**

Odo Benda is specialised in the development of XML database applications for diverse international and national research projects (e.g. REGNET, MEDIA.ALP, MODOK, DISMARC, DIS). Currently he is working on the implementation of thesaurus Web services for the cultural heritage domain (based on international standards such as ANSI and SKOS vocabulary). Within DISMARC and DIS he is responsible for the thesaurus and vocabulary implementation for the virtual catalogues.

**Mag. Gerda Koch** – Managing Director of AIT Ltd.

Gerda Koch collaborated in MOSAIC for the Austrian partner of the project. Her work focuses on cultural digital content provisions (e.g. for MODOK, COVAX). 2001-03 she was responsible for the financial and administrative coordination of the EU IST-Project REGNET (23 partners in 12 European states), coordinating also the legal framework for the project continuation. For DISMARC and DIS she works on the metadata and vocabulary mapping, user interface presentation, system validation and coordinating regional management issues.

### **9.1.6 Atos Origin System Integration (ATOS)**

Atos Origin is a leading international IT services provider. We provide integrated design, build and operate solutions to large multi-national clients in carefully targeted industry sectors. Its business is turning client vision into results through the application of consulting, systems integration and managed operations. The company's annual revenues are EUR 5.4 billion and it employs 50,000 people in 40 countries. Atos Origin is the Worldwide Information Technology Partner for the Olympic Games and has a client base of international blue-chip companies across all sectors.

At Atos Origin, Systems Integration is not just about integrating new solutions, but includes getting the most out of legacy applications to prolong returns from existing IT investment. Successfully combining new solutions with established ones can transform the complete enterprise architecture into a single, seamless business system. Our extensive experience in integrating people, processes and technologies enables us to design, build and operate practical and robust solutions.

**Roger Essoh** – Atos Public Sector Business Developer and Head of innovation

In charge of relation with FEDORA Commons, the preservation and archive open source foundation. Driving several innovations in digital content preservation and archive, member of the Stanford University "Enterprise Repositories and Federated Archives" working group. Managing the R&D team for the French National Library Preservation and Archive Distributed System (SPAR).

**Maxime Ametti** – Project manager of SPAR

Engineering diploma (Ecole Centrale de Lille) and Pre-doctoral research in industrial data processing. Coordinates the IT team (about 12 people). Takes part to the steering committee

meetings and tracking meetings. Manages vendors and external expertise. In charge of the reporting to the Atos Origin management.

### **Lee Namba**

Technical Architect. Master's of Electrical Engineering - Cornell University New York, USA and Bachelor's of Science Electrical Engineering - University of California San Diego, USA. Defines the global architecture of SPAR based on FEDORA. Sets up the development and integration environments (software configuration management, continuous integration, development frameworks such as Spring, etc.) and best practices.

### **Gautier Poupeau** – SPAR functional consultant

Ecole nationale supérieure des sciences de l'information et des bibliothèques (ENSSIB). Defines all the rules to manage the data handled by BnF. Expertise in Semantic Web, XML, OAIS and specific schemas (such as METS, TextMD, MIX, etc. defined by the Library of Congress).

### **Charlotte Fabre** – SPAR functional leader

Ecole nationale supérieure des sciences de l'information et des bibliothèques (ENSSIB). Expertise in OAIS and preservation of data. Defines the processes for the preservation of data in accordance with OAIS standard. Works with the development team to check that the system external specifications are correctly handled.

### **Laurent Sollier** – SPAR technical leader

Engineering diploma. Drives the development team, helps the technical architect in the definition of the system. In charge of the system internal specifications.

## **9.1.7 Freie Universität Berlin (FUB-BGBM)**

The Botanic Garden and Botanical Museum Berlin-Dahlem (BGBM), with its extensive scientific collections of herbarium specimens (about 3.5 million) and living plants, is a centre of biodiversity research in Europe. It houses the most complete botanical library in the Germany. The library holds a wide range of literature on plants from all over the world, in all printed languages and from five centuries, among them many precious and very rare books. BGBM recognised early the new role of natural history museums in the domain of electronic information. Today, the BGBM has a separate department of Biodiversity Informatics with, at present, 20 staff members (information scientists, botanists, zoologists, engineers, mathematicians, and technicians). Focal point of research and development activities are taxonomic information systems and networking of distributed primary biodiversity information. The IT infrastructure includes 26 Linux and Microsoft servers, with PostgreSQL, MySQL, and SQL-Server database management systems. Broadband network capacity is available via a Gigabit backbone as well as a Gigabit connection to the European Scientific network GÉANT via GWIN.

### **Prof. Dr. Walter Berendsohn**

Director of Dept. of Biodiversity Informatics and Laboratories of BGBM. 1983 to 1990 at the Jardín Botánico La Laguna in El Salvador, from 1987 as research director. 1998 - head of the Dept. of Biodiversity Informatics and Laboratories, 2000 - appointed as Director and Professor. His efforts have been primarily devoted to information systems in biodiversity research. He has authored several information models and coordinated numerous EU projects or project work packages with many partners. He has acted as coordinator for federal programs in biodiversity informatics research in Germany, and is a member of the German delegation to the GBIF governing board. He also chairs the coordinators' group of GBIF Germany, committees in the TDWG Biodiversity Information Standards organisation, and the GBIF Science Subcommittee for the Digitisation of Collections.



### **Anton Güntsch**

MSc in Computer Science, head of the Biodiversity Informatics and Documentation section in the Dept. of Biodiversity Informatics and Laboratories of BGBM. Scientific activities: design and implementation of collection and taxonomic databases at the meta and object levels; design of cooperative networks of distributed biodiversity information systems; Digitisation of living and conserved biological collections. Implementation of the BioCISE collection catalogue and the European Natural History Specimen Information Network Pilot system, an XML-based World Wide Web access system for distributed heterogeneous collection databases. Present project and committee memberships include: CODATA Working Group on Biological Collection Data Access, GBIF Germany IT commission (chair), Euro+Med computer working group, and EDIT Information Science and Technology Commission.

**Dr. Norbert Kilian** – Head of the section Library, Archives and Publications of BGBM

Diploma in Biology, PhD in Botany (Taxonomy). His activities include the scientific supervision of the library, the editorship of the BGBM's international scientific journal "Willdenowia" and research in the field of taxonomy and systematics of vascular plants. A particular focus of his research activities is to explore the possibilities of the Web for the cooperative production and user-friendly presentation of taxonomy. He is coordinator of one taxonomic group in the EU-funded EDIT project.

#### **9.1.8 Georg-August-Universität Göttingen Stiftung Öffentlichen Rechts (AnimalBase) (UGOE)**

The EZOOLO/AnimalBase project is located at the Georg-August-Universität of Göttingen. The university was founded in 1737 and established since then a great international reputation both in the natural sciences and in the arts. The EZOOLO/AnimalBase project was initiated as a joint venture of Göttingen University Library (SUB) and the Zoological Institute of the university to provide free access to digitised versions of all taxonomically relevant early zoological works. Since 2003 more than 400 monographs and numerous journals were digitised. The names described in these works were extracted and transferred to AnimalBase ([www.animalbase.org](http://www.animalbase.org)). In a first step (2003-2005) more than 100,000 pages were digitised from the earliest beginnings of scientific zoology in the 1550s until the year 1770, and 10,000 animal names were entered to AnimalBase. In a second step from 2008 onwards the literature shall be covered until the 1820s, with approximately 50,000 animal names.

#### **Dr. Francisco Welter-Schultes**

Francisco Welter-Schultes has worked since 1992 as a zoologist specialised in European non-marine mollusc faunistics, taxonomy and systematics in the Zoological Institute at UGOE and holds a PhD in this field. His research projects included compiling bibliographies for mollusc faunas in SE Europe, in the course of which he gained profound experience with early zoological literature. In 2003 he initiated the EZOOLO/AnimalBase project.

#### **9.1.9 Naturhistorisches Museum Wien (NHMW)**

The Natural History Museum in Vienna was founded more than 250 years ago. The museum includes departments of Zoology, Botany, Geology-Palaeontology, Mineralogy, Anthropology and Karst and Caves. Most departments are organised in "collections" for special taxonomic units with responsible scientific curators. The collections with more than 30 million specimens including hundreds of thousands of types are the basis for any taxonomic work. The specimens are also the basis for many investigations in ecology or biodiversity. AT-TAF (SYNTHEsys) is one of the "first addresses" for taxonomic work. Additionally a library with many historically important volumes is available. The library with c. 6000 scientific journals and tens of thousands of books complement with the National Library of Austria and the University's libraries nearby. Therefore, many old and rare books are available for taxonomic work.

**Dr. Ernst Vitek** – Interim Director of Department of Botany

Besides being the Interim Director, he is also the Head curator of the collections of phanerogams. Since 1997 he is the editor of the "Annalen des Naturhistorischen Museums, Serie B". Together with his colleagues, he is member of the Austrian working group for long-term storage.

**Heimo Rainer**

He is a Botanist and Zoologist and strongly involved in the TDWG (Biodiversity Information Standards). Since 2000 he is responsible for database management and digitisation within a project aiming in the digitisation of the herbaria of the University of Vienna and the NHMW.

**9.1.10 Oberösterreichisches Landesmuseen (OOELM)**

The Biology Centre in Linz-Dornach, with its more than 6 million objects, represents the largest natural history collection in the province of Upper Austria and is the 2nd largest in Austria. It currently publishes the series *Stapfia* and *Denisia* and three other journals. It holds the biodiversity database ZOBODAT, founded in 1972 as ZOODAT. The database today includes more than 3.3 million records (~38,000 species) on the distribution of animals and plants. Furthermore we have included literature citations (more than 33,000), OCR scanned books (~150,000 pages) and bibliographies from about 4,000 biologists until now.

**Dipl.-Ing. Michael Malicky** – Head of IT at OOELM

Michael Malicky started his career as IT Administrator of the Biodiversity Database ZOBODAT in 1991 at Linz University. In 1998 he graduated as Dipl.-Ing. for Informatics with a special interest in databases. He has participated in the following projects: 2002-2005 EU Project ENBI, Work package Coordinator; 2002-2006 Sp2000 Europa, Content Provider; since 2001: several projects within GBIF Austria; since 2006: Project "Digitisation of Upper Austrian and Austrian Natural History Literature" with F. Gusenleitner.

**Mag. Fritz Gusenleitner** – Deputy Director of the Biology Centre at OOELM

Mag. Fritz Gusenleitner is curator for Entomology at OOELM since 1981. He is also editor and chief editor of the following series: *Linzer Biologische Beiträge*, *Beiträge zur Naturkunde Oberösterreichs*, *Entomofauna*, *Enomologica Austriaca*, *Stapfia*, *Denisia*.

**9.1.11 Hungarian Natural History Museum (HNHM)**

HNHM is one of largest natural history museums in Central-Eastern Europe. More than 10 million natural history items are preserved in the collections. The collections hold important reference and historical information especially for the Carpathian Basin, the Balkans, Asia Minor, Central and Inner Asia, Eastern Asia, Africa and South America. HNHM Library contains more than 300,000 volumes. The catalogue of the library is available on the Internet. HNHM plays an important role as publisher of natural history. Several natural history journals and books published by the HNHM during its 200 years of history. During the last years these are available via the internet but there is a strong commitment by HNHM to digitise and provide free access to its own journals and books.

**Angéla Matuszka** – Head of Library and Science History Collection at HNHM

She was the editor of *Officina Nova Publishers* (2002-2003); editor of *Bastei* (2001-2002) and from 1995 to 2001 information retriever and editor of CO-NEX Ltd. Ms. Matuszka holds a degree in history and library from the L. Eötvös University in Budapest.

**László Peregovits** – Chief Museologist, Department of Zoology

He joined the HNHM in 1983 as a research biologist and has several decades of experience in editing and publishing including conceptual and technical aspects. He was involved in various research projects relevant to the present proposal: Digitisation of Insects of Mongolia (GBIF), BioCASE, FaEu (Fauna Europe), SYNTHESYS, EDIT. Since 2002 he is member of CETAF, representative of HNHM.

**Dr. Miklós Rajczy** – Head of Information Technology Group

He joined the HNHM in 1976 as the Curator of Bryophytes. He has been involved in various projects, including: MUSoNET, Digitisation of HNHM mushroom herbarium data, Digitisation of HNHM herbarium data, Database of Hungarian published data on algae. He is a committee member of EDIT, the Hungarian Biodiversity Platform, and the Advisory Committee to the Hungarian Biodiversity Monitoring System.

**9.1.12 Museum and Institute of Zoology, Polish Academy of Sciences (MIZPAS)**

The origin of the Library of the Museum and Institute of Zoology dates back to the 19th century. After establishing the National Museum of Natural History in 1919, the Library received numerous donations which formed its basis. The collection is of national importance with holdings of zoology, especially systematic and zoogeography, entomology and ornithology. Recently it has increased by publications concerning genetics and molecular biology. At present the Library comprises 243,271 volumes and 5,378 archival items (books 44,876 vol.; reprints 74,065 vol.; journals 116,760 vol.; 6,071 serial titles; 696 current serial titles; maps 4,953 units; voices 134 units).

**Robert Turlej** – EU Programs Coordinator

Robert Turlej holds a Master degree in Banking and Insurance. Before joining the MIZPAN as PL-TAF (SYNTHESYS) and EDIT coordinator he worked as Sales Director for Qui-Tech Sp. z.o.o. and IT expert for Com-Tech.

**Monika Malcher** – Library Manager

Monika Malcher holds a Masters degree of the Faculty of Animal Sciences of Warsaw University. She is managing the library activities of MIZPAN since 2000.

**9.1.13 The Natural History Museum of Denmark, University of Copenhagen (SNM)**

The museum holds an estimated 12 million specimens of animals, plants, books, archives, fossils, minerals, and other natural history related items. It is situated in the Botanical Garden, which is also part of the museum. Part of the museum is three libraries, a botanical, a zoological, and a geological. The botanical library was founded in 1752 and holds 154,000 bibliographic entities (books, journals, reprints). The zoological library has a close connection to the Science library, which is a part of the national library. The geological library holds 125,000 bibliographic entities. The museum hosts the GBIF and the Danish GBIF node.

**Henning Knudsen**

He is mycologist, curator of fungi and collections manager for SNM incl. all collections except the garden, but including the libraries and research leader for the botanists. Main research in agarics in arctic and temperate areas on the northern Hemisphere, main work as editor and co-author of Nordic Macromycetes and the forthcoming Funga Nordica. Has taken the initiative to digitise important Danish botanical works like Flora Danica and Flora Agaricina Danica, both finished, but only the second online so far.

**Christian Lange** – IT-coordinator at SNM

He is responsible for digitisation and the database programming in SNM, for photographing, for running our scanning equipment etc. He is also responsible for the homepage of the Danish mycological society.

**9.1.14 Stichting Nationaal Natuurhistorisch Museum Naturalis (NAT)**

Naturalis was founded in 1820 and much of its collection dates back to the 19th and 20th century. Naturalis has a staff of about 160 people, which include scientists, collection managers, exhibition designers, information officers, educators, etc. The collections of zoological, palaeontological, and geological objects are estimated about 12 million objects. With reference to the BHL-Europe

project Naturalis can rely on a strong and innovative department of information services backed by natural history collections and archives which cover nearly 200 years of research and collecting.

**Kees Hendriks** – Head of Information Services

Kees Hendriks studied medicine at the University of Leiden. Before starting to work at the National museum of Natural History Naturalis in 2000 he worked for various employers as project assistant, project manager, department manager and vice director. The focus in all the jobs fulfilled is information exchange. Kees Hendriks now works as acting deputy director of Public Engagement. And is responsible among other things for the policy development in science communication both geared to popular audiences as well as to scientific target groups.

**Tom Gilissen** – Information professional

Tom Gilissen has a degree in Information Sciences and works for the library of Naturalis.

**Berry van der Hoorn** – Project Manager Information Services

Berry van der Hoorn studied biology and worked for nine years as a project manager and advisor at the largest insurance company of the Netherlands, where he was responsible for the development and implementation of several Web based information systems. Most of his projects concern the development of Web sites and Web based applications.

**Dr. René W R J Dekker** – Director of Collections

With a staff of about 35 collection and data managers René Dekker is responsible for the policy and management of the collection.

### 9.1.15 National Botanic Garden of Belgium (NBGB)

NBGB is a ‘complete’ botanical garden, integrating a living collection (‘Hortus’; 18,000 species in cultivation) and a large museum (‘Herbarium’; more than 3 million specimens plants incl. fungi), thus aiming at research, conservation and education. Research in taxonomy, phylogeny and allied fields covers all plant groups and geographically focuses on Western Europe and tropical Africa. NBGB maintains an extensive scientific library. Particular emphasis is placed on systematic botany, floristics, plant ecology, botanical history, ethnobotany and horticulture and this concerning all systematic groups of non-vascular plants and Fungi, as well as vascular plants. The library holds 50,000 monographs including 2,500 valuable historical books (the oldest from 1486), 5,000 periodicals and 25,000 reprints. Historical literature on Central African flora was kept in NBGB since 1890.

**Relevant expertise:** Data repatriation to partners both in Africa and Latin America is a priority for NBGB. The institute is the leader internationally with regard to central African floristics, and keeps the majority of the botanical material ever collected in RD Congo, Rwanda and Burundi. Drawings and colour paintings of flowering plants and fungi were digitised and linked to specimens kept in the BR herbarium. NBGB is an institutional member of CETAF and is active in the EC projects ENBI, EDIT, and SYNTHESYS. It also collaborates as a training provider to GTI (Global Taxonomy Initiative) capacity building activities with developing countries.

**Prof. Dr. Elmar Robbrecht** – Head of Department ‘Vascular plants’

Elmar Robbrecht has been working at NBGB since 1974. He holds a MSc and PhD in Botany from the State University of Ghent (Belgium). Since 1994, he is a Member of the Belgian *Royal Academy for Overseas Sciences* (section Natural and Medical Sciences; vice-president in 2000, president of the section in 2001).

**Nicole Hanquart**

Nicole Hanquart holds a MSc in History and is presently doing a master degree in library management at the Free University of Brussels. She has been working at the Library of NBG since 1996 and is presently curator of the collection of valuable books. She was involved in the

databasing of the whole library collection of NBGB and is in charge of the management of the scientific archives of the institution.

### **Régine Fabri**

Régine Fabri holds a PhD in Botany, and a MSc in library management from the State University of Liège (Belgium). She has been working in NBGB since 1982. In 1994 she was transferred to the library as responsible of the computerising of the catalogue. Since 2004, she is scientific responsible of the library and is now also in charge of the library administration.

### **9.1.16 Royal Museum for Central Africa (RMCA)**

RMCA is a multidisciplinary institution with a special focus on Sub-Saharan Africa. It is a leading research institute and knowledge centre on the biodiversity of living species in the context of their natural environments in Africa and aims to develop interest and understanding for African fauna in the scientific communities and the public at large. The museum manages collections of about 10 million specimens of animals and 56,000 wood specimens of 13,600 different botanical species. The RMCA maintains an extensive library on African biodiversity, including the top scientific journals, but also a unique collection of rare, old colonial publications. Scientific staff speak ten languages; consequently, the library and reprint collection is unusually multilingual. The institution has about 1.2 km of Archives.

**Relevant experience:** Information on biodiversity is repatriated to African partners, towards which also training in taxonomy is provided. RMCA is an institutional member of TDWG (Biodiversity Information Standards) and involved in Biodiversity Informatics initiatives like GBIF, ENBI, EDIT, CETAF, SYNTHESYS, FishBase, EOL and CBOL (Consortium Barcode of Life). It collaborates to GTI capacity building activities with developing countries.

**Michel Louette** – Head of Dept. & Scientific Director of the Dept. for African Zoology

Michel Louette has been working at the RMCA since 1974. During 1982-83 he holds a Fulbright/Hays Fellowship when he was visiting professor at the Sacramento State University, CA, U.S.A. From 1985 to 2000 he was a Member of the Pan African Ornithological Congress Committee, from 1992 to 1996 it's Chairman. Since 1990 he was Member of the International Ornithological Committee and since 2006 Member of the Standing Committee on Ornithological Nomenclature of the IOC.

### **Dr. Patricia Mergen**

Patricia Mergen is responsible for external relations and project management at RMCA since 2005. She holds a PhD in Biological Sciences from University of Namur in Belgium. From 2002 to 2005 she was project manager of the Belgian Biodiversity Information Facility (BeBIF). She then served as an independent expert for the evaluation of FP6 proposals. Currently she is President of the Royal Belgian Zoology Association (2007-2008). Other memberships include the GBIF ICT Advisory Group, WP13 of ENBI, and TDWG. From 2002 to 2005 she was a member of the Belgian Delegate of the GBIF Nodes Committee and the GBIF DADI subcommittee.

Additionally the project will benefit from input of the biodiversity information unit (Danny Meirte, Bart Meganck, Garin Cael, Franck Theeten and Kim Jacobsen), from the staff of RMCA's Metafro-Infosys, our History department and Central library.

### **9.1.17 Royal Belgian Institute of Natural Sciences (RBINS)**

Founded in 1846, the RBINS houses a diverse and exceptionally rich zoological collection, palaeoanthropology, prehistoric items and a diverse mineral collection. The total is in the order of 37,000,000 specimens with around 100,000 primary types. This places RBINS among the world top ten collections in terms of volume of specimens stored. The scientific library of the RBINS is the biggest documentary resource of natural history in Belgium. It offers a vast range of books (695,368 volumes) and has very specialised, often unique scientific magazines. Its catalogue is available

online. RBINS has started the digitisation of the catalogue of the library and more than 185,000 titles are online on the RBINS Web site.

**Dr. Patrick Grootaert** – Head of Department of Entomology at RBINS

Patrick Grootaert studied Zoology at the State University of Ghent Belgium and holds a PhD with a specialisation in Nematology. He also obtained a Doctor in Science degree in Entomology. He is curator of the Diptera collection and published more than 450 scientific papers. As BE-TAF manager he ensures the links between the National Botanical Garden, the Royal Museum for Central Africa and RBINS. Since 28 June 2008 he became vice-director of his Institute keeping his capacity of research and curator of Entomology.

**Laurent Meese** – Head of the Library at RBINS

Laurent Meese obtained a master degree in Communication Sciences and later specialised in library sciences. He coordinates the digitisation of the catalogue of the library as well as the pilot project of the digitisation of the Dautzenberg reprint collection. He is involved in the Driver2 programme of digitisation coordinated by the University of Ghent.

### 9.1.18 Bibliothèque nationale de France (BnF)

BnF is one of the largest public and research libraries in the world and holds more than 50,000 monographs and around 3,000 titles of periodicals published in the natural sciences field between 1801 and 1920. The BnF offers access to its digital library Gallica ([gallica.bnf.fr](http://gallica.bnf.fr)), created through the library's commitment to the digitisation of selected items of its collections. More than 100,000 printed volumes digitised, in French and other languages are freely available online. From the origins of printing to the early 20<sup>th</sup> century, they cover all domains of knowledge, with a specific focus in literature and history. Since 2007, BnF has been improving its digitisation programmes, with 100,000 new items being digitised each year, in text and image mode. In spring 2008, the BnF launched a new version of Gallica with new and modern functionalities drawing upon the most recent Web 2.0 experience. At a European level, the BnF is a founding member of The European Library consortium and is involved in the TELplus and IMPACT projects where it is looking at high quality OCR, full-text indexing and subject multilingual issues. The BnF is also a member of the EDLnet thematic network.

**Eva Fuentes** – Head of the Unit «Sciences de la vie, sciences de l'ingénieur»

She is in charge of the management of this collection and in charge of its digitisation.

**Hervé Colinmaire** – Director of the Science and Technology Department of BnF

With a staff of about 100 people, he is responsible for the management and “valorisation” of the collections. He has a degree in Information Sciences. Prior to the BnF, he worked as a director in several public and university libraries.

**Frédérique Joannic-Seta** – Head of the Unit “Associated Centres-Gallica”

She is in charge of the management of the digital library Gallica as well as the coordination of the cooperation with the French libraries and organisations (all centres of excellence in their specific fields) in order to implement a policy of sharing and complementing document holdings and digitisation. She has a degree in Information Sciences and in Archives.

### 9.1.19 Museum national d'histoire naturelle (MNHN)

The mission of MNHN, which was founded in 1793, is ‘to discover, understand, highlight and help to preserve Earth's natural and cultural diversity’. Its main activities are research, education and training, enrichment of its collections (around 68 million specimens), providing expertise, and diffusion of scientific knowledge. MNHN was and is heavily involved in various European programmes related to biodiversity: The European Topic Centre on Biological Diversity of the European Environmental Agency; the French GBIF, GTI and CBD nodes, the French Focal Point

for the EU-funded BioCASE project; leading partner in several past and on-going EU projects, including ParSyst, ColParSyst, Fauna Europaea, ENBI and EuroCAT/Sp2000 Europa, SYNTHESYS, EDIT, MarBEF, EUMon, and PESI; branch of the FISHBASE consortium; hosting several global databases of taxonomic references all linked to international initiatives such as GBIF, Sp2000, EOL; a member of the Consortium for Barcode of Life (CBOL). All these actions are also supported by a central Library Department, which holds the world's third collection of literature, original drawings and manuscripts relating to natural history. The print collections include 20,000 periodical titles and 600,000 books with the oldest dating from 15<sup>th</sup> century. Digitisation plans are already on-going in full collaboration with BnF (mussi.mnhn.fr).

### **Mrs Michelle Lenoir**

Mrs Michelle Lenoir graduated the National High School for Information Sciences and Librarianship in 1971. Since 2003 she is the Head Librarian of the MNHN, the national reference library in natural history, where she is charged with implementing the Museum collections development policy. Previously in her career, she played a key role as an expert in heritage collections for the French Ministry of Culture.

### **9.1.20 Consejo Superior de Investigaciones Científicas (CSIC)**

CSIC is the largest National Research Institution in Spain. The participant institute, Museo Nacional de Ciencias Naturales (MNCN) houses the biggest natural history collections, library and archives in Spain. It comprises the most competitive research groups in Spain working on biodiversity systematics, biogeographic modelling, conservation biology, palaeobiology and geology, among others. Currently, staff members lead or take an active part in 221 projects, including 10 funded by the EU. Relevant national project for the proposal lead by MNCN is Fauna Ibérica. MNCN is founding member of CETAF, partner of SYNTHESYS, EDIT, LifeWatch and coordinate the Spanish GBIF node. The public library of MNCN has the most complete coverage in the Iberian Peninsula for natural history literature, both in terms of classical and recent books, and CDs. It contains more than 62,000 volumes and more than 6,400 scientific journals as well as access to more than 9,000 electronic journals. This library offer online access to the main bibliographic collective catalogue of CSIC in Spain (CIRBIC), and to other numerous catalogues and databases of natural history publications as well as to a wide range of scientific journals.

### **Antonio G. Valdecasas**

Antonio G. Valdecasas is a senior researcher at the MNCN. He has worked in computer applications to taxonomy, building the first truly hierarchical taxonomist database (1989) and the first Web page of a Spanish Museum (1994). Other developments are in image processing for taxonomy. In 2006 he has developed the first iconic online database for water mites.

### **Marian Ramos – Vice-Director for Research at MNCN**

Marian Ramos heads the Fauna Ibérica project, the edition of its monographs and its online database IBERFAUNA, is ES-TAF Leader (SYNTHESYS), former CETAF Chair, the Spanish Focal Point for the GTI, Chair of the Bern Convention Group of Experts on Invertebrates (Council of Europe) and WP leader of EDIT and of LifeWatch.

### **Isabel Morón – Head of the Library at MNCN**

Isabel Morón has participated in numerous national digitisation and diffusion projects of bibliographic holdings. Regular trainer in cataloguing courses. She has published articles on the holdings of the MNCN library.

### **9.1.21 Museo di Storia Naturale dell'Università degli Studi di Firenze (MSN)**

The Natural History Museum of Florence was founded in 1775 by Grand Duke Peter Leopold. With more than 10 million specimens, it is the most important natural history museum in Italy. The Museum houses specimens of extraordinary scientific and naturalistic value: from the XVI century

herbaria to the valuable XVIII century waxes, from the fossil elephant skeletons to the collections of brightly coloured butterflies, from the huge tourmaline crystals to the Aztec artefacts, from the imposing wooden sculptures to the world's largest flower. A lot of ancient and rare books are also preserved in the Library (Biblioteca di Scienze).

#### **Luca Bartolozzi** – Head of Zoology

Luca Bartolozzi holds a MSc in Biological Sciences from the University of Florence. He is Curator of the Entomology Department and is a specialist in tropical insect biodiversity (systematics, taxonomy, and faunistics of Lucanidae and Brentidae (Coleoptera)). He has published more than 100 scientific papers.

#### **9.1.22 Royal Botanic Garden Edinburgh (RBGE)**

RBGE is an internationally renowned centre for botanical research and conservation work and holds one of the largest collections of living plant species in Europe, together with extensive collections of preserved plant and fungal material. It has one of the most important botanical libraries in the UK and has a fully functioning and highly productive molecular laboratory. At present the RBGE has 32 researchers, 9 curatorial, 4 technicians and 6 librarians.

#### **Jane Hutcheon** – Head Librarian Royal Botanic Garden Edinburgh

A chartered librarian with over 30 years experience in academic and research librarianship. External committees: Immediate past president European Botanical & Horticultural Libraries Group. Member of the Council for Botanical & Horticultural Libraries. Member of Scottish Agricultural Libraries Group. Member of Rare Books in Scotland Group. External Assessor Wellcome Trust History of Medicine Research Resources in Medicine. Validation Panel Member for Edinburgh's Telford College Library and Information Science Course. Supervisor for postgraduate Library and Information Science Students. Panel Member for the Integrated Collections Strategy, National Library of Scotland. Site Supervisor for National Association for Decorative and Fine Arts volunteer.

#### **9.1.23 Species 2000 (Sp2000)**

Species 2000 is a Network organisation that creates an index of the world's known organisms. Its distributed model synthesises the index from sectors supplied by taxonomic databases across Europe and around the world, many from the major European institutions of CETAF. The programme reached production scale as an EC scientific infrastructure under the FP5 EuroCAT project, and as a member of EDIT celebrated coverage of one million species in 2007. Its Catalogue of Life is a global service ([www.catalogueoflife.org](http://www.catalogueoflife.org)) recognised by the UN Convention on Biological Diversity, and presently comprising a synonymic species checklist of 1.1 million plants, animals, fungi and micro-organisms, about 2.5 million names, and a comprehensive taxonomic hierarchy. It provides a taxonomic backbone for global biodiversity portals, such as EOL and GBIF, for about 40 national portals worldwide, and is used by scientists in 79 countries. It contributes content and the taxonomic hierarchy used by the uBio taxonomically intelligent tool in the BHL programme.

#### **Prof. Dr. Frank Bisby**

Frank Bisby is a founding member of TDWG, instigator of one of the first global species databases, the International Legume Database and Information Service (ILDIS), and leader of the Species 2000 organisation since its launch in 1996, as well as member/work package leader/co-coordinator in EC projects such as ERMS, Euro+Med, ENBI, EuroCAT and EDIT. As Executive Director of Species 2000 he works with what is now an extensive network of taxonomic database hubs around the world from China and New Zealand to Brazil and N. America, in projects to approach completion of known organism coverage in the Catalogue of Life, and to work with partner organisations on the Global Names Architecture.



**Dr. Yuri Roskov**

Yuri Roskov is a plant taxonomist and completed his PhD on the genus *Trifolium*. He subsequently led the ILDIS Northern Eurasia database programme. Since 2002 he worked at the University of Reading on taxonomic information products in projects such as ILDIS, ENBI, and EuroCAT, and is now Executive Editor of the Catalogue of Life.

**9.1.24 John Wiley & Sons limited (Wiley)**

Wiley is a global publishing company founded in 1807 that markets its products to professionals and consumers, students and instructors in higher education, and researchers and practitioners in scientific, technical, medical, and scholarly fields. The company produces books, journals, and encyclopedias, in print and electronically, as well as online products and services, training materials, and educational materials for undergraduate, graduate, and continuing education students. Through the 2007 acquisition of Blackwell Publishing, Wiley has gained the Blackwell Synergy platform home to over 850 Blackwell journals. Also during 2007 the company completed an initiative to digitise its entire historical journal holdings, making 8.2 million pages of content dating back to 1799 available on Wiley InterScience.

**Philippa Scoones** – Web Publishing Director

**Liz Ferguson** – Associate Publishing Director

**9.1.25 Smithsonian Institution (SIL)**

The Smithsonian Natural History Museum is dedicated to inspiring curiosity, discovery, and learning about the natural world through its unparalleled research, collections, libraries, exhibitions, and education programs. At the centre of the Museum's exhibition and research programs are its expertly documented collections: more than 125 million natural science specimens and cultural artefacts including 30 million insects, 4.5 million plants, 7 million fish, and 2 million cultural artefacts. Over 3.5 million specimens are out on loan each year; over 15,000 visitor days are spent in the collections; and there are almost 600,000 additional visits to collection data bases available on the Web. The BHL is led from the Smithsonian.

**Thomas Garnett** – Program Director of the BHL

Thomas Garnett was Associate Director for Digital Library and Information Systems at SIL. He has over 20 years experience creating, scoping, implementing, and managing major digital library projects. He is interested in networking biological data, literature, institutions, and projects to enable investigation of complex systems.

**9.1.26 Missouri Botanical Garden (MOBOT)**

The mission of MOBOT is to discover and share knowledge about plants and their environment, in order to preserve and enrich life. Founded by Henry Shaw and opened to the public in 1859, the Garden is a National Historic Landmark and widely considered one of the top three botanical gardens in the world. The Garden's goals include achieving excellence in the areas of horticultural cultivation and display; global botanical research, cataloguing, information dissemination, and conservation; science and environmental education; home gardening education; and public outreach. MOBOT is a founding member of the BHL and is supporting the development of the system's infrastructure, application layers, and interfaces.

**Chris Freeland** – Director of Bioinformatics at MOBOT & Technical Director of BHL

Chris Freeland an M.S. in Biological Sciences from Eastern Illinois University and has worked as a technologist in scientific organisations for more than 10 years. He has been a project manager for several large informatics projects, including the development of the Tropicos botanical information system, online at [www.tropicos.org](http://www.tropicos.org), and the BHL.

## 10 Appendices

### 10.1 Financing Plan

Applicant Short Name	Source of funding (Amounts in euros)				
	Contribution from own resources	Contribution by other organisation*	Direct revenues expected from the project**	Contribution requested from the Commission	Total
UBER	115,640 EUR	0 EUR	0 EUR	462,560 EUR	578,200.00 EUR
NHM	219,290 EUR	0 EUR	0 EUR	877,160 EUR	1,096,451 EUR
NMP	39,367 EUR	0 EUR	0 EUR	157,466 EUR	196,833 EUR
EDL	38,340 EUR	0 EUR	0 EUR	153,360 EUR	191,700 EUR
AIT	74,398 EUR	0 EUR	0 EUR	297,594 EUR	371,992 EUR
ATOS	223,068 EUR	0 EUR	0 EUR	892,272 EUR	1,115,340 EUR
FUB-BGBM	39,405 EUR	0 EUR	0 EUR	157,622 EUR	197,028 EUR
UGOE	16,150 EUR	0 EUR	0 EUR	64,600 EUR	80,750 EUR
NHMW	19,570 EUR	0 EUR	0 EUR	78,280 EUR	97,850 EUR
OOELM	14,402 EUR	0 EUR	0 EUR	57,608 EUR	72,010 EUR
HNHM	5,774 EUR	0 EUR	0 EUR	23,096 EUR	28,870 EUR
MIZPAS	6,590 EUR	0 EUR	0 EUR	26,360 EUR	32,950 EUR
SNM	8,790 EUR	0 EUR	0 EUR	35,160 EUR	43,950 EUR
NAT	13,300 EUR	0 EUR	0 EUR	53,200 EUR	66,500 EUR
NBGB	14,149 EUR	0 EUR	0 EUR	56,598 EUR	70,747 EUR
RMCA	18,640 EUR	0 EUR	0 EUR	74,560 EUR	93,200 EUR
RBINS	13,860 EUR	0 EUR	0 EUR	55,440 EUR	69,300 EUR
BnF	17,516 EUR	0 EUR	0 EUR	70,064 EUR	87,580 EUR
MNHN	14,960 EUR	0 EUR	0 EUR	59,840 EUR	74,800 EUR
CSIC	12,893 EUR	0 EUR	0 EUR	51,571 EUR	64,463 EUR
MSN	3,770 EUR	0 EUR	0 EUR	15,080 EUR	18,850 EUR
RBGE	10,129 EUR	0 EUR	0 EUR	40,518 EUR	50,647 EUR
Sp2000	6,907 EUR	0 EUR	0 EUR	27,627 EUR	34,534 EUR
Wiley	1,740 EUR	0 EUR	0 EUR	6,960 EUR	8,700 EUR
SIL	0 EUR	0 EUR	0 EUR	0 EUR	0 EUR
MOBOT	0 EUR	0 EUR	0 EUR	0 EUR	0 EUR
<b>TOTAL</b>	<b>948.649 EUR</b>	<b>0 EUR</b>	<b>0 EUR</b>	<b>3.794.596 EUR</b>	<b>4.743.245 EUR</b>

Applicant Short Name	*Details on contributions by other organisations	**Details on direct revenues expected from the project
All (AP 1-26)	Not relevant	Not relevant

## 10.2 Other specific costs

### Applicant 1 (UBER):

The Kick-Off Meeting and the Final Review meeting will take place in Berlin (Museum für Naturkunde). The considerable number of applicants and people involved require a proper event management including catering and equipment. The costs given below are based on experience with similar events.

Kick-Off Meeting	15,000 EUR
Final Review	15,000 EUR
<b>Total</b>	<b>30,000 EUR</b>

### Applicant 2 (NHM):

The principle hardware and proprietary software components related to this project are the provision of the centralised aggregated storage of scanned images and related metadata together with the provision of Web services and downloadable tool sets. This will require a dedicated storage solution, tape library, operating software, backup management software, and servers to host Web services etc. The estimated cost break down for these components is:

Storage Solution	500,000 EUR
Tape Library	100,000 EUR
Servers	45,000 EUR
Operating/backup software	70,000 EUR
Commissioning	80,000 EUR
<b>Total</b>	<b>795,000 EUR</b>

### Applicant 3 (NMP):

Some expenses will be incurred in relation to the effective implementation of the communication activities. These expenses are related to Web site hosting and publication of the BHL-Europe promotion kit:

Web site hosting	500 EUR
Web site design	7,000 EUR
Printing (e.g. brochures, posters, leaflets, postcards, gadgets)	7,000 EUR
Translation	5,000 EUR
BHL-Europe journal	4,000 EUR
Instruments to support newcomers to join	2,500 EUR
<b>Total</b>	<b>26,000 EUR</b>

### 10.3 List of events and meetings

#### List of Events and Meetings

Meeting	Date (Project month)	Participants	Location
Kick-off meeting	Month 1	All partners	Berlin
Technology Workshop	Month 1	Technology providers	Berlin
WP2 Meeting	Month 3	Content providers	London
Technology Workshop	Month 3	Technology Management Board	London
PMG Meeting	Month 3	PMG	London
MoU-Agreement	Month 6	Technology + Content providers	Prague
Dissemination Workshop	Month 6	Communication Working Group	Prague
PMG + PSG meeting	Month 6	PMG + PSG	Prague
PMG Meeting	Month 9	PMG	Berlin
PMG + PSG meeting	Month 12	PMG + PSG	Vienna
PMG + PSG meeting	Month 18	PMG + PSG	Leiden
Dissemination Workshop	Month 24	Communication Working Group	Paris
Technology Workshop	Month 24	Technology Management Board	Paris
PMG + PSG meeting	Month 24	PMG + PSG	Paris
PMG + PSG meeting	Month 30	PMG + PSG	Copenhagen
Final review	Month 36	All partners	Berlin

## 10.4 Background and reference documents

### 10.4.1 Biodiversity Heritage Library Workplan – English Language Component

#### **BHL Work Plan: Building a Digital Open Access Library for Biodiversity Literature**

In February 2004, biologists, librarians, and information specialists gathered for a four-day workshop sponsored by the Pinhead Institute in Telluride, Colorado to assess the feasibility of assembling a Web-based Encyclopedia of Life (EOL). Among the outcomes of this workshop was a recommendation to digitise the literature of biodiversity. In 2005, an international symposium, *Library and Laboratory: the marriage of research, data, and taxonomic literature*, funded by the Alfred P Sloan Foundation, was held at the Natural History Museum in London. Over 80 biologists, librarians, and computer scientists attended. Participants identified the lack of access to the published literature of biodiversity as one of the principal obstacles to efficient and productive research. Few scientific disciplines are as dependent on the historical literature in their respective field. In the spring of 2005, representatives of ten major natural history museum libraries, botanical libraries, and research institutions joined in a collaborative effort to develop strategies to digitise the literature in an open access manner. From this partnership grew the Biodiversity Heritage Library (BHL) project.

The partners envision that a research scientist or student who has access to the Internet, located anywhere in the world, will be able to search for specific information in all of the literature relevant to biodiversity and transparently link the documentation to relevant taxonomic, geographic, or other useful databases. Such a tool would erase much of the expensive, labour-intensive work of library research and speed the production of research results many times over.

The greatest diversity of the biota exists in tropical and developing countries, yet the literature documenting that biodiversity is primarily held in a few North American and European libraries. Digitising this literature and making it available freely on the Internet will be an act of significant intellectual repatriation. Currently, teachers and scientists from developing countries who need to read this literature must visit major, expensive urban centers such as London, New York, or Washington, D.C. using very limited funds. All BHL libraries report remote visitors forming lines at their copying machines to bring needed material back to their home countries.

The BHL partnership is essential because while natural history museum and botanical garden libraries have collected biodiversity materials comprehensively, including many specialised and rare materials, no single library holds the complete corpus of legacy literature. The partners' collections represent a uniquely comprehensive assemblage of this literature. Within two years of the start of this project, the BHL will provide approximately 15 million digitised pages of literature to support multiple bioinformatics initiatives and research. For the first time in history, the core of our natural history museum and botanical garden libraries will be available to a truly global audience.

#### **Methodology**

BHL member institutions will scan (or have already scanned) their institution's own scientific publications to contribute to the BHL. Members will scan other volumes from their collections that are not covered by copyright or for which permissions have been obtained using the Internet Archive, a non-profit partner, to quickly add large segments of literature to the corpus. Ongoing negotiations with commercial publishers and learned society publishers are promising: newer literature may be made available with permissions. The BHL as a community-based partnership should provide a trusted grouping to negotiate with copyright owners. Small society publishers need help scanning and storing their publications and if we provide this service, some are pleased to have their content accessible through the BHL portal. A set of clearly defined working relationships will be established as part of the project. User input to the process will be solicited regularly.

The BHL selected the Internet Archive (IA), an organisation with demonstrated technical qualifications for mass scanning and long-term digital content management, to scan the bulk of the literature through scanning centers. A scanning center consists of multiple, high-speed, state-of-the-art digital book scanners, with staff for two shifts daily, each able to handle large numbers of volumes daily. The IA will perform imaging, OCR, association of standard metadata (derived from MARC records provided from the libraries' catalogs) with the digitised files, file arrangement, security, maintenance of optimal book conservation practices and delivery of the completed scans in conformance with agreed project standards. Article-level access to journals will be provided. In many cases BHL libraries will partner with other local libraries to support Internet Archive scanning centers in Washington, D.C., New York, Boston, London, and elsewhere.

### **BHL Portal with Taxonomic Intelligence**

The digitised literature will be served to users from the Biodiversity Heritage Library Portal to be hosted initially by the Missouri Botanical Garden. The BHL Portal will create an innovative research environment that will vastly accelerate research in life sciences and conservation: a freely accessible, service-based Web Service formed through coupling existing databases with digitised, searchable images and OCR text of heritage literature. The BHL Portal will use existing informatics tools to identify strengths and overlap across the participating institutions' libraries and to help solve the problems associated with the naming of organisms over time.

### **Taxonomic Intelligence**

The digitisation of a major corpus of biodiversity literature will advance world biodiversity initiatives significantly, but only to the extent that users can find relevant content. Names of organisms annotate content about species. However, the use of names for information retrieval is impeded because names are neither stable nor consistent. One organism may have more than one name. This prevents simple automated indexing services from bringing together complementary data. Moreover, about 1% of names change each year, such that the many-names-for-one-organism (synonyms) problem accumulate with time and will be particularly severe with heritage literature. Visitors to traditional library scanning projects who know organisms by their colloquial (common) names may be unable to find content unless they know the names used in the source documents. These issues will reduce the utility of the millions of pages of primary biodiversity information to be generated by the BHL without the added tools intended in the BHL Portal. The uBio team from the Marine Biological Laboratory / Woods Hole Oceanographic Institution (MBLWHOI) library has assembled an array of taxonomically intelligent services designed to overcome these problems.

### **Selection of Materials to be Scanned**

Because BHL requires moving physical objects (books) and contracting for expensive scanning machines that are built to order and require extensive set up in specific locations, it is necessary that selection of locations and material be done with clear priorities and reasons. BHL involves 10 separate institutions with different collection strengths. Selection of material to be prioritised will be multi-dimensional including thematic areas, date of publication, and quantity of material. BHL Directors have organised a "BHL Collections Working Group", which will refine and further articulate the thematic areas in June and July. The initial titles to be scanned will either be in the public domain or have copyright permission. Input from the EOL Secretariat will be the initial focus of the themes assigned. In addition to thematic focus provided by the EOL Secretariat, the BHL will also analyse such major indexes as *Index Kewensis*, Sherbourne's *Index Animalium*, and Neave's *Nomenclator Zoologicus*. BHL is in discussions to obtain permissions to mine *Zoological Record* to determine those journals that have been most cited in the literature of species identification and description. This will provide a priority list of journals and monographs. Scanning from a prioritised list created through citation analysis ensures that the BHL contains the most critical works for our audience of scientists and scholars. Additional thematic focus will be possible post-scanning after applying taxonomic intelligence to draw out all the species names, so that a filter can be run with, for example, the species names from the Ocean Biogeographic Information System (OBIS) in order

to tag those articles or monographs with marine data. Other name sources can be employed in a similar manner. BHL will also remain flexible enough to take timely advantage of offers from significant learned society journals to digitise back holdings in an open access manner.

As the EOL Secretariat establishes thematic areas of emphasis for the EOL, the BHL libraries will select texts that best support the thematic area. Initial thematic assignments are subject to revision in light of the Secretariat's priorities, for 2007-2008.

MBL and Harvard MCZ will prioritise marine literature and expect to digitise 70% of the pre-1923 marine literature. Smithsonian will initially prioritise entomological literature and expects to complete 40% of the pre-1923 literature. Missouri Botanical Garden, the New York Botanical Gardens, Harvard Botany, and the Royal Botanic Gardens, Kew will prioritise botany and expect to digitise 70% of the pre-1923 literature.

#### **10.4.2 ViTaL – The Virtual Taxonomic Library in the EU EDIT Programme**

ViTaL is part of the EDIT project, which is creating a European Distributed Institute of Taxonomy. EDIT's goal is to create a network supporting taxonomy for biodiversity and ecosystem research. EDIT is creating tools and forging links between institutions throughout Europe and beyond which will help reduce many of the impediments currently affecting biological taxonomic work. ViTaL's part in this is to aid the discovery and accessibility of taxonomically relevant literature through the provision of four main services:

- ViTaL's metasearch component will allow searching across a targeted set of taxonomic resources, including library catalogues, databases and e-journals.
- ViTaL will make available an OpenURL resolver, allowing users to click through their search results to full text content where this is available (through resources such as e-journals or repositories like the Biodiversity Heritage Library), or otherwise to a set of links providing search functions.
- ViTaL will include a bibliographic reference aggregator, which will harvest references from a range of sources and to provide a centralised bibliographic resource.
- ViTaL will also provide a facility which will enable users of taxonomic literature to nominate suitable materials for digitisation. These requests will be available to users of the site and also to members of digitisation projects, allowing them to make use of this information in the prioritisation and planning of scanning schedules.

The components making up ViTaL are currently under development at the Natural History Museum, London. ViTaL is being developed by staff at the Museum in consultation with the EDIT applications development team based in Berlin at Botanical Gardens (FUB-BGBM).

#### **10.4.3 Researchers comments concerning biodiversity literature digitisation**

- 1) Currently we at PhilRice have a global taxonomic project on the rice black bug (*Scotinophara* spp.). Please we need to refer to the ancient scientific literature on this insect. At the National Library in Philippines, we do not have access to all of the literature listed in the attachment (attached file). Please can you help us on how to get access to them from your library. (*Dr. Ravi Josh, Chief Scientist at the Philippine Rice Research Institute*)
- 2) First of all congratulations for the Botanicus.org Webpage, it's a very useful project specially if you are located at countries far away from decent libraries (as I am in Costa Rica). I like the PDF version of the books, as it can be downloaded; it is easy to read parts of the books off-line. (*Walter Schug*)

- 3) Thank you so much for not only your quick response to my ILL request, but even more for your attaching the item as a PDF file so that Prof. Newton received it almost instantly across the ether. I know he emailed you a much more timely thank you. He wrote me that he is extremely excited about your digitisation project. At the moment he and his graduate botany students in Kenya have access to very few resources. He spends his summer terms at Kew doing his research for the next year's teaching and writing, but he tells me that now, because of what is already on your site, he will not have to carry so much back to Kenya for his research and his students but can download and work with your resources right there. (*Emilie Pulver, University of Kenya*)
- 4) I am absolutely amazed of this tool! I think it is fantastic what we can offer to the botanical community and beyond to have at our finger tips. Congratulations to you and the staff involved. Thanks for all your good work. (*Carmen Ulloa*)
- 5) In reference to: *Bulletin des Séances de la Société Entomologique de France*. [Paris]: Société entomologique de France, [1873-1884]: My deepest gratitude for allowing me access to the digital version of the very rare "Bulletin des Séances de la Société Entomologique de France". It has been very important for my work on the database of the names of the butterflies of the world to be able to consult at leisure this series, which is held by extremely few libraries in the world. I cannot stress enough the importance of having access to electronic versions of the literature, especially to us researchers who cannot benefit from well-endowed institutional libraries. The Smithsonian Libraries are doing a great service to science by making openly accessible such crucial works as the "Biologia Centrali-Americana", and now the above-mentioned "Bulletin". I only wish that there were many more such electronic resources. Please keep up the excellent work! (*Dr Gerardo Lamas, Museo de Historia Natural, Universidad Nacional Mayor de San Marcos*)
- 6) In reference to: Frederick Ducane Godman and Osbert Salvin, eds. *Biologia Centrali-Americana*. [London: Pub. for the editors by R. H. Porter], 1879-1915: I have to my position the collection of Coleoptera of the Faculty of Superior Studies of the Independent National University of Mexico and daily I consult volumes of BCA for the identification of specimens of Coleoptera, because it is a wonderful work and until the moment does not exist another source that supports in the identification of the Mexican species of several families of this group. (*Ma. Magdalena Ordóñez Reséndiz, Museo de Zoología, FES Zaragoza, UNAM*)
- 7) In reference to: Walter Rothschild. *The Avifauna of Laysan and the neighbouring islands with a complete history to date of the birds of the Hawaiian possession*. London: R H Porter, 1893-1900: Aloha. I live on The Big Island of Hawai'i, a \$300.00 plane ride away from Honolulu and the Bishop Museum. Even when I can make it to the Museum (where I study the Hawaiian Bird Skins), they do not have every single bird (moho apicalis, the Oahu moho is missing)...I have been looking for this text for over TWENTY YEARS. Mahalo nui loa for all your hard work. Reading these pages mean so much to me and many others. I hope they show there appreciation as well. It truly is very important. I cannot thank you enough, nor stress the importance of your Web site enough. Thank you for putting these items on the Web, and in such a findable manner. (*Gwendolyn O'Connor*)
- 8) In reference to: Howard Jones. *Illustrations of the nests and eggs of birds of Ohio*. Circleville, Ohio, 1886: Virginia Hunt, a doctoral candidate at LSU, will be using this site as part of her dissertation on ornithological illustration; commented Ms. Hunt: "My study concerns surveying a large number of ornithological narrative paintings, in particular historical examples, in order to determine how these may be used by high school and college biology teachers to teach certain key concepts in ecology while doing 'double duty' in illustrating subtle aspects of the history and nature of science. (*Bruce Shelvey, Ph.D.; Chair, Department of Geography, History, and Political and International Studies and Associate Professor of History and Political Studies, Trinity Western University, Canada*).



- 9) In reference to: The Journal of the Bombay Natural History Society: This is perhaps the greatest resource for those interested in the biodiversity of the Indian region. This availability and hopefully more material is a great advance for humankind. The founders of the BNHS started the organisation for the sharing of information - something that the Internet does in a way that would make the founders truly happy. This material which would be under public domain under the Indian Copyright Act 1957 has hitherto not been available except to the very select few members who hold membership in the club, live in Bombay and have the interest. Some other copies in other libraries around India have either been maintained in a poor state, eaten by dermestids, crumbled over time or have been vandalised by people. (*Shyamal*)